



Premier League

English Premier League Predictions

Darion Ong Yu Jie

TABLE OF CONTENTS

01 INTRODUCTION

02 PROBLEM STATEMENT

03 THE DATA APPROACH

04 EXPLORATORY DATA ANALYSIS &
FEATURE ENGINEERING

05 MODEL SELECTION & EVALUATION

06 PREDICTIONS

07 CONCLUSION



INTRODUCTION

Football is the most popular sport in the world and it is very dominant in countries from Europe, South America and Africa.

Each country have their own individual leagues and English Premier League is the top league from England.

Notable teams in league are Manchester United, Liverpool, Arsenal, Chelsea and Manchester City.



PROBLEM STATEMENT

Using past matches stats
and football team stats
from FIFA , predict future
matches results.



Approach :

Using ML with Multi-Class Classification, discover important features for predicting the match results and develop a model which can assist potential stakeholders such as football pundits, betting website, football fans and shareholders of football club.

THE DATA APPROACH



Web Scrap Data for the
following seasons:
2017/2018, 2018/2019
2019/2020, 2020/2021

Chosen players' stats are
then taken from the FIFA
stats to form team stats.

EXPLORATORY DATA ANALYSIS & FEATURE ENGINEERING

Team Statistics Features

- As mentioned in previous slide, the team stats are engineered from top 18 players with the highest appearances in the season.
- The stats used from each player are Value, Wages, Age, Height, Weight, Overall, Potential, International Reputation and Skills Move.
- Different features are engineered differently:
 - Wage and Value are summed.
 - Mean and Standard Deviation of Age, Height, Weight, Overall, Potential, International Reputation and Skills Move are used

club	total_value_eur	total_wage_eur	average_age	average_height_cm	average_weight_kg	average_overall
Arsenal	285300000	1287000	25.24	182.12	76.35	78.76
Aston Villa	141370000	751000	24.89	183.00	75.22	74.00
Brighton & Hove Albion	124025000	693000	24.94	182.06	76.53	74.18
Burnley	137700000	776000	27.59	182.88	78.06	75.53
Chelsea	437500000	1673000	25.18	181.94	78.53	81.59

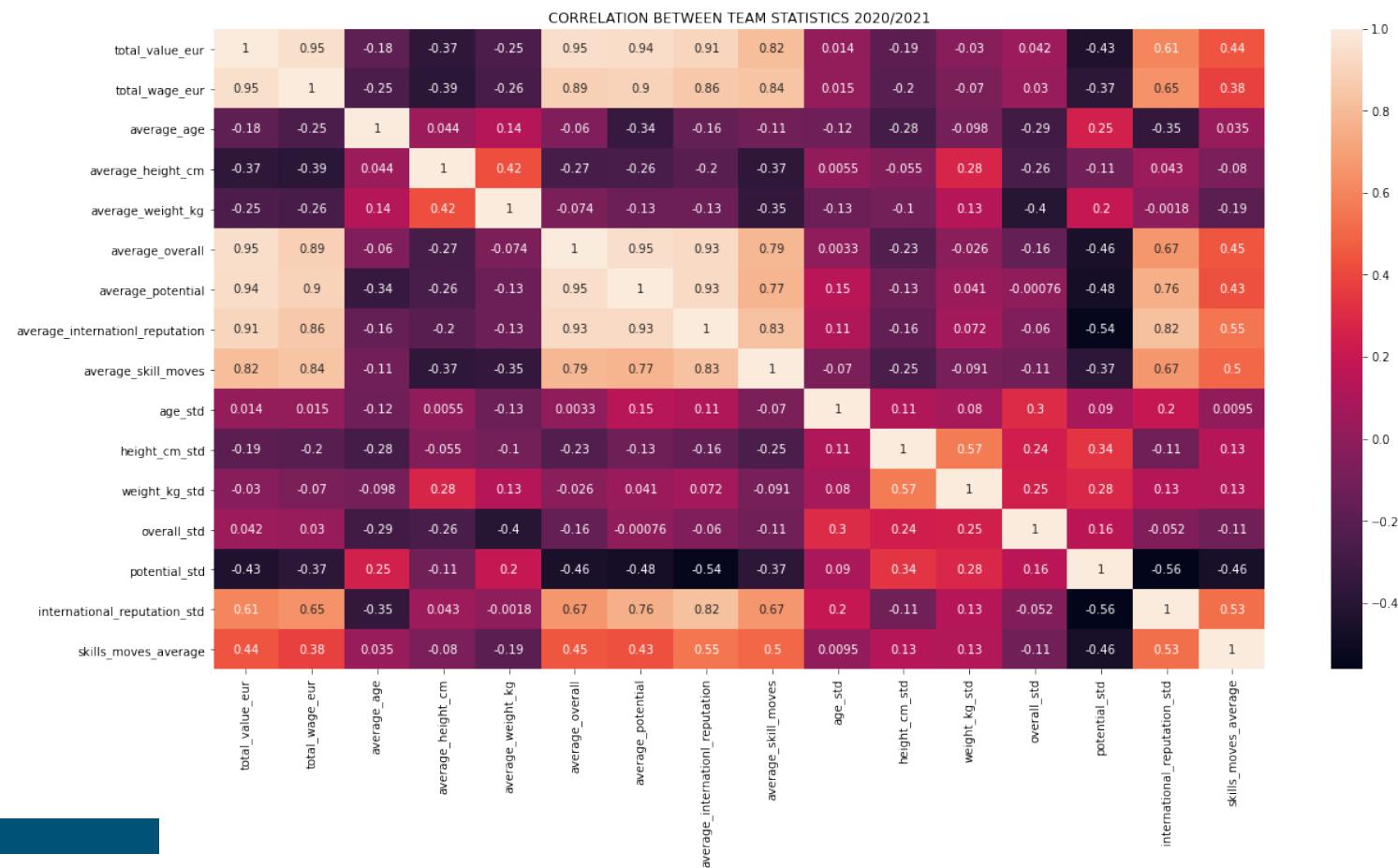
average_potential	average_internationl_reputation	average_skill_moves	age_std	height_cm_std	weight_kg_std	overall_std
83.82	1.94	3.12	4.22	5.56	5.96	4.45
79.00	1.22	2.94	2.93	5.86	8.80	5.21
78.76	1.29	2.65	3.58	8.79	7.37	4.50
77.24	1.12	2.65	3.37	6.12	6.59	4.50
85.59	2.29	3.06	4.63	6.95	8.46	3.61

potential_std	international_reputation_std	skills_moves_std
2.81	1.09	0.86
3.24	0.55	0.80
2.99	0.59	1.06
3.70	0.33	0.79
2.90	0.99	0.90

EXPLORATORY DATA ANALYSIS & FEATURE ENGINEERING

Correlation Between Team Statistics

- This is the heatmap of correlation between all the team stats.
- As footballers tend to be more valuable if they are better, the average overall and average potential are positive correlated with value and wages.
- However, as a footballer ages, they tend to be not as good as before, their value, potential, overall and wages will drop.
- Therefore, age has a strong negative correlation with a lot of other features.



EXPLORATORY DATA ANALYSIS & FEATURE ENGINEERING

Fixtures Matches Stats

- One of the datasets we are using is the fixtures data, which consists of the matches that have been played in the season.
- Each match has the match stats for the players, such as Gls which is goals, SoT is shots on target and CrdY is yellow card.
- As we are looking at the team as a whole, we will sum up each stat from each player into one.

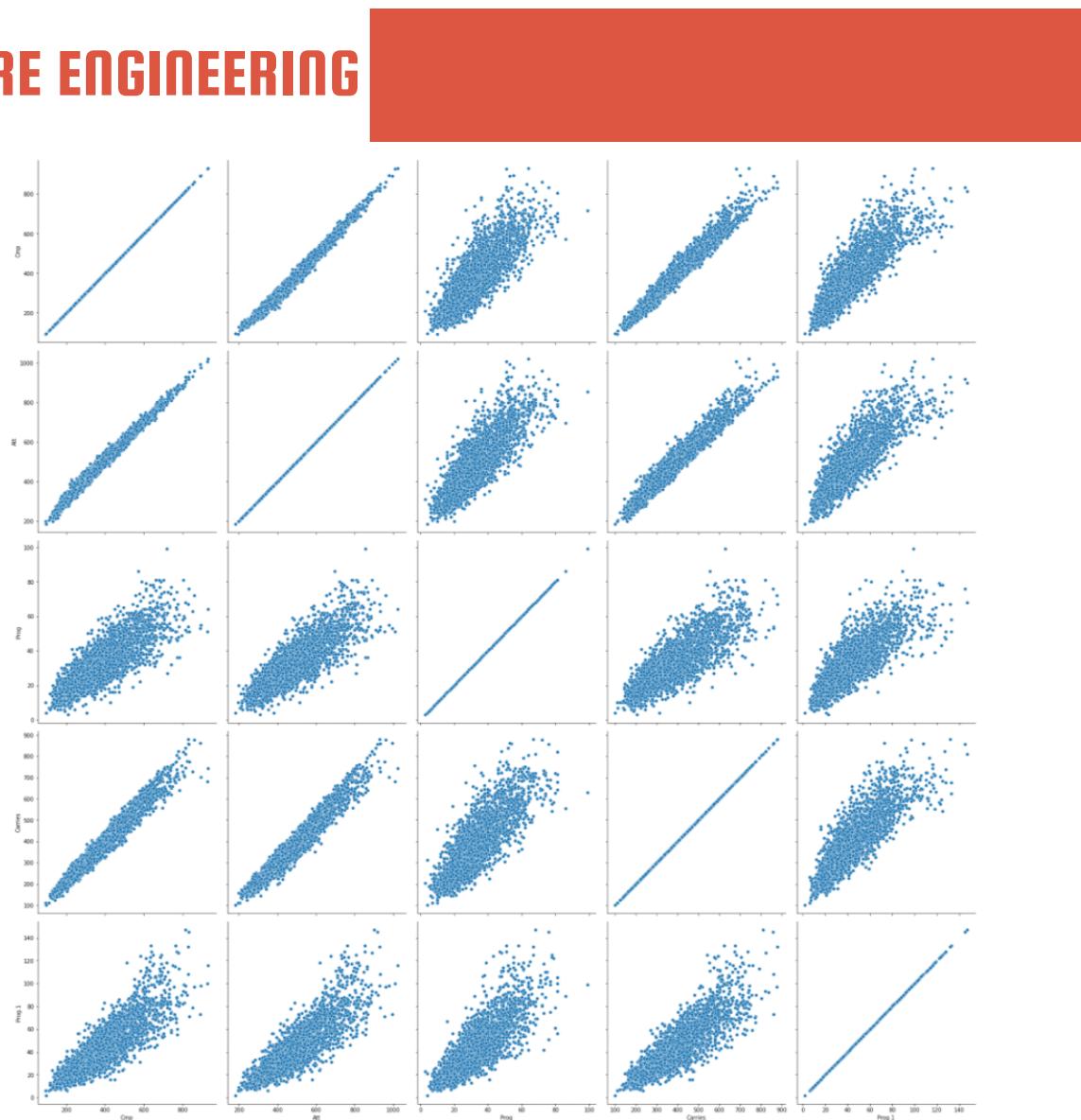
Player	#	Nation	Pos	Age	Min	Gls	Ast	PK	PKatt	Sh	SoT	CrdY	CrdR	Touches	Press	Tkl	Int_x	Blocks
Alexandre Lacazette	9.0	fr FRA	FW	26-104	75.0	1	0	0	0	2	2	0	0	22.0	26.0	1.0	1	0.0
Olivier Giroud	12.0	fr FRA	FW	30-344	15.0	0	0	0	0	1	1	0	0	14.0	4.0	0.0	1	0.0
Danny Welbeck	23.0	eng ENG	AM	26-287	75.0	2	1	0	0	3	2	0	0	36.0	21.0	0.0	0	4.0
Alexis Sánchez	7.0	cl CHI	AM	28-264	15.0	0	0	0	0	2	1	0	0	13.0	4.0	0.0	0	0.0
Mesut Ozil	11.0	de GER	AM	28-329	90.0	0	0	0	0	2	1	0	0	65.0	33.0	2.0	0	1.0

Match	Home	Away	Stadium	Attendance	Gls	Ast	PK	PKatt	Sh	SoT	CrdY	CrdR	Touches	Press	Tkl	Int_x	Blocks
Arsenal_vs_Bournemouth	Arsenal	Bournemouth	Emirates Stadium, London	59262	3	3	0	0	16	8	0	0	717.0	169.0	18.0	19	8.0
Arsenal_vs_Bournemouth	Bournemouth	Arsenal	Emirates Stadium, London	59262	0	0	0	0	8	2	1	0	560.0	225.0	8.0	23	10.0

EXPLORATORY DATA ANALYSIS & FEATURE ENGINEERING

Fixtures Matches Stats

- Most of the features in the match stats are related.
- The grid on the right contains the scatter plot for:
 - Passes Completed (Cmp) > 1st Row
 - Passes Attempted (Att) > 2nd Row
 - Progressive Passes (Prog) > 3rd Row
 - Times the team controlled the ball (Carries) > 4th Row
 - Progressive Carries (Prog.1) > 5th Row
- Using the scatter plot, you can see the moderate to strong linear relationship between the features. These features have strong relationship.
- Based on my understanding, these features are related to passes and possession. Before a pass is carried out, the ball will be at the player's possession.



EXPLORATORY DATA ANALYSIS & FEATURE ENGINEERING

Past Matches Stats

- As we will not know future matches stats as the match will not been played, we will engineer the last three games and last five games average to determine the future.

Match	last_3_avrg_Gls	last_5_avrg_Gls	last_3_avrg_Sh	last_5_avrg_Fls	last_3_avrg_Fld	last_5_avrg_PKwon
West_Bromwich_Albion_vs_Leicester_City	0.666667	1.0	10.000000	15.0	10.000000	0.0
Norwich_City_vs_Watford	0.333333	0.4	9.000000	13.6	11.000000	0.0
Burnley_vs_Arsenal	1.333333	1.4	6.666667	14.8	14.333333	0.0
Manchester_City_vs_Chelsea	1.666667	1.4	15.000000	10.0	11.000000	0.2

EXPLORATORY DATA ANALYSIS & FEATURE ENGINEERING

Result & Match Goals

- There is no result feature.
Therefore, create using the goals
and own goals stats.
- The values will be as followed: 0 is
Home Win, 1 is Draw and 2 is Away
Win.
- The home goals feature does not
include the away own goals and is
the same for the away goal
feature.
- Three new features were created,
Home Total Goals, Away Total
Goals and Total Goals.

	match	home	h_gls	h_og	a_gls	a_og	result	h_total_goals	a_total_goals	total_goals
0	Arsenal_vs_Leicester_City	Arsenal	4	0	3	0	0	4	3	7
1	Crystal_Palace_vs_Huddersfield_Town	Crystal Palace	0	1	2	0	2	0	3	3
2	Chelsea_vs_Burnley	Chelsea	2	0	3	0	2	2	3	5
3	Brighton_&_Hove_Albion_vs_Manchester_City	Brighton & Hove Albion	0	1	1	0	2	0	2	2
4	West_Bromwich_Albion_vs_Bournemouth	West Bromwich Albion	1	0	0	0	0	1	0	1
5	Everton_vs_Stoke_City	Everton	1	0	0	0	0	1	0	1
6	Watford_vs_Liverpool	Watford	3	0	3	0	1	3	3	6

EXPLORATORY DATA ANALYSIS & FEATURE ENGINEERING

Combination of Fixtures and Team Stats

- After all necessary features are sorted, we combined the fixtures and team stats accordingly to the seasons they played in.
- Each season is ranged between two dates based on the table on the right
- Season 2020/2021 is currently ongoing.

	Start Date	End Date
Season 2017/2018	2017-08-09	2018-05-13
Season 2018/2019	2018-08-09	2019-05-13
Season 2019/2020	2019-08-08	2020-07-27
Season 2020/2021	2020-09-11	Current

MODEL SELECTION & EVALUATION

With our data, we will be looking at three targets.

1. Result
2. Home Total Goals
3. Away Total Goals



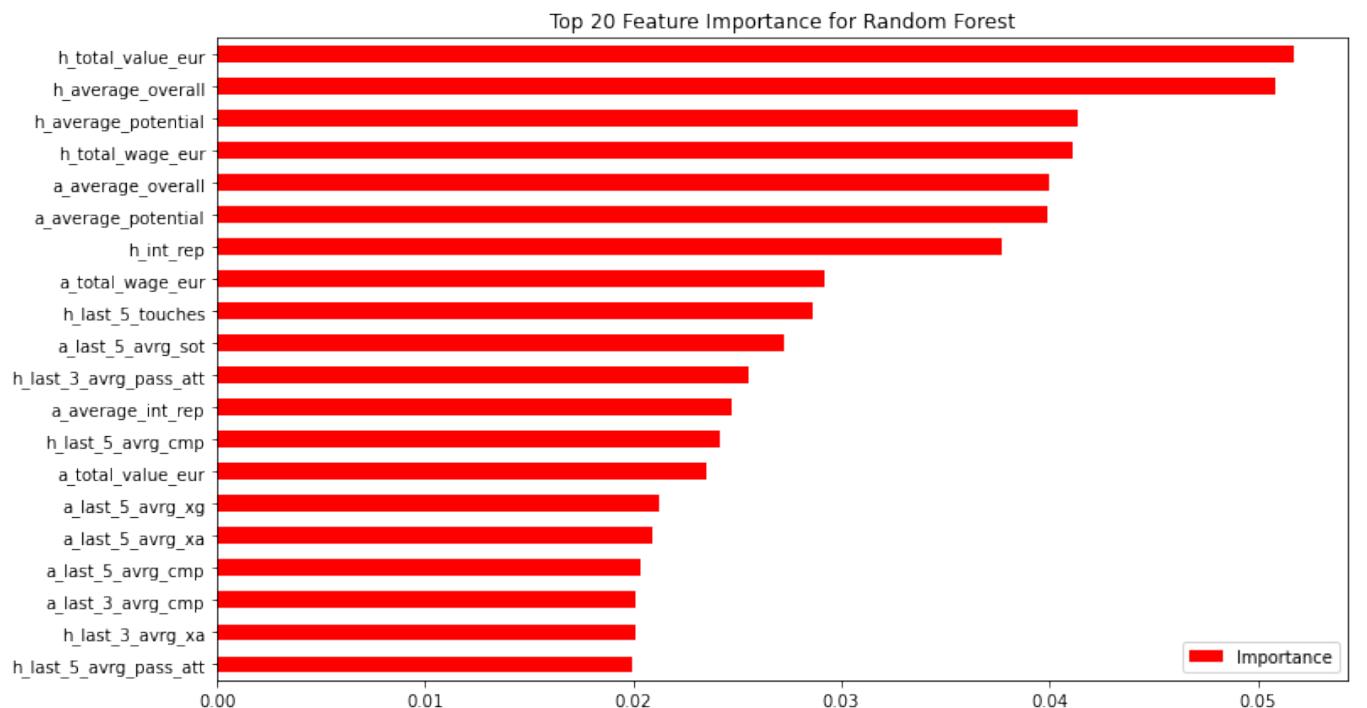
MODEL SELECTION & EVALUATION: Result

- As we prepare for modeling, we have dropped rows with null values and features that are not so related to Result.
- F1 Score is used as the determinant metric. Micro Average is used as the data is not balanced.
- Random Forest is chosen.

	Model Type	F1-Score	Recall	Precision	Train Accuracy	Test Accuracy
1	Random Forest	0.534351	0.534351	0.534351	0.618197	0.534351
0	Logistic Regression	0.526718	0.526718	0.526718	0.564626	0.526718
5	Support Vector Classifier	0.519084	0.519084	0.519084	0.582483	0.519084
2	Extra Trees	0.511450	0.511450	0.511450	0.570578	0.511450
3	AdaBoost	0.496183	0.496183	0.496183	0.606293	0.496183
4	GradientBoosting	0.496183	0.496183	0.496183	0.601190	0.496183

MODEL SELECTION & EVALUATION: Result

- Using Random Forest, most of the features that are important are individual team stats.



MODEL SELECTION & EVALUATION: Home Total Goals

- Similarly, we will drop null values and features not very correlated to Home Total Goals. There are a total of 9 classes. However, we remove 2 of them as there are only two matches in it.
- The data is still very imbalanced, so they are SMOTE before being trained.
- F1 Score is used as the determinant metric. Micro Average is used as the data is not balanced. However, the test accuracy is lower than the base line.
- Gradient Boosting is chosen.

	Model Type	F1-Score	Recall	Precision	Train Accuracy	Test Accuracy
4	GradientBoosting	0.312883	0.312883	0.312883	0.987692	0.312883
3	AdaBoost	0.300613	0.300613	0.300613	0.465055	0.300613
5	Support Vector Classifier	0.285276	0.285276	0.285276	0.802637	0.285276
1	Random Forest	0.279141	0.279141	0.279141	0.639560	0.279141
0	Logistic Regression	0.254601	0.254601	0.254601	0.507253	0.254601
2	Extra Trees	0.245399	0.245399	0.245399	0.599560	0.245399

MODEL SELECTION & EVALUATION: Away Total Goals

- Similarly, we will drop null values and features not very correlated to Away Total Goals. There are also a total of 9 classes. However, we remove 2 of them as there are only two matches in it.
- The data is still very imbalanced, so they are SMOTE before being trained.
- F1 Score is used as the determinant metric. Micro Average is used as the data is not balanced. However, the test accuracy is lower than the base line.
- Support Vector Classifier is chosen.

	Model Type	F1-Score	Recall	Precision	Train Accuracy	Test Accuracy
5	Support Vector Classifier	0.322086	0.322086	0.322086	0.706398	0.322086
4	GradientBoosting	0.315951	0.315951	0.315951	0.994303	0.315951
3	AdaBoost	0.306748	0.306748	0.306748	0.315074	0.306748
2	Extra Trees	0.303681	0.303681	0.303681	0.688431	0.303681
1	Random Forest	0.300613	0.300613	0.300613	0.713409	0.300613
0	Logistic Regression	0.263804	0.263804	0.263804	0.534181	0.263804

PREDICTIONS:

Result Odds with Singapore Pools

- To test the model, we decide to take matches that are to be played this week for predictions.
- We use the current season dataset to determine the upcoming matches.
- On the right is the table of matches with the predicted results and probabilities.
- As some matches have just been played, the below table is the result.

	date	match	home	away	predicted result	home_win	draw	away_win
0	2021-01-26	Newcastle_United_vs_Leeds_United	Newcastle United	Leeds United	2	0.285313	0.304331	0.410356
1	2021-01-26	Crystal_Palace_vs_West_Ham_United	Crystal Palace	West Ham United	0	0.365430	0.271783	0.362788
2	2021-01-26	West_Bromwich_Albion_vs_Manchester_City	West Bromwich Albion	Manchester City	2	0.158648	0.105513	0.735839
3	2021-01-26	Southampton_vs_Arsenal	Southampton	Arsenal	2	0.312912	0.224345	0.462744
4	2021-01-27	Burnley_vs_Aston_Villa	Burnley	Aston Villa	2	0.345117	0.279015	0.375868
5	2021-01-27	Everton_vs_Leicester_City	Everton	Leicester City	0	0.432772	0.252232	0.314996
6	2021-01-27	Chelsea_vs_Wolverhampton_Wanderers	Chelsea	Wolverhampton Wanderers	0	0.642730	0.228803	0.128467
7	2021-01-27	Manchester_United_vs_Sheffield_United	Manchester United	Sheffield United	0	0.720578	0.194367	0.085056
8	2021-01-27	Brighton_&_Hove_Albia_vs_Fulham	Brighton & Hove Albion	Fulham	0	0.441512	0.267817	0.290671
9	2021-01-28	Tottenham_Hotspur_vs_Liverpool	Tottenham Hotspur	Liverpool	1	0.316149	0.371714	0.312137

ENGLAND - PREMIER LEAGUE					JANUARY 27
FT	Crystal Palace	2 - 3	West Ham United	X	★
FT	Newcastle United	1 - 2	Leeds United	✓	★
FT	Southampton	1 - 3	Arsenal	✓	★
FT	West Bromwich Albion	0 - 5	Manchester City	✓	★

PREDICTIONS:

Result Odds with Singapore Pools

- Using probability, we managed to calculate the odds created from our model.
- We then compared the odds to the odds by Singapore Pools. This is to see whether we think alike.

date	match	home	away	predicted result	home_win	draw	away_win
0	2021-01-26 Newcastle_United_vs_Leeds_United	Newcastle United	Leeds United	2	3.50	3.29	2.44
1	2021-01-26 Crystal_Palace_vs_West_Ham_United	Crystal Palace	West Ham United	0	2.74	3.68	2.76
2	2021-01-26 West_Bromwich_Albion_vs_Manchester_City	West Bromwich Albion	Manchester City	2	6.30	9.48	1.36
3	2021-01-26 Southampton_vs_Arsenal	Southampton	Arsenal	2	3.20	4.46	2.16
4	2021-01-27 Burnley_vs_Aston_Villa	Burnley	Aston Villa	2	2.90	3.58	2.66
5	2021-01-27 Everton_vs_Leicester_City	Everton	Leicester City	0	2.31	3.96	3.17
6	2021-01-27 Chelsea_vs_Wolverhampton_Wanderers	Chelsea	Wolverhampton Wanderers	0	1.56	4.37	7.78
7	2021-01-27 Manchester_United_vs_Sheffield_United	Manchester United	Sheffield United	0	1.39	5.14	11.76
8	2021-01-27 Brighton_&_Hove_Albia_vs_Fulham	Brighton & Hove Albion	Fulham	0	2.26	3.73	3.44
9	2021-01-28 Tottenham_Hotspur_vs_Liverpool	Tottenham Hotspur	Liverpool	1	3.16	2.69	3.20

Wed, 27 Jan 2021							
Time	Event	Home (1)	Draw (X)	Away (2)			
2.00am	0304 Crystal Palace vs West Ham	01 2.80	02 3.00	03 2.30	[M]	[⌚]	[(+11)]
2.00am	0360 Newcastle vs Leeds	01 3.40	02 3.50	03 1.85	[M]	[⌚]	[(+11)]
4.15am	0359 Southampton vs Arsenal	01 2.95	02 3.10	03 2.15	[M]	[⌚]	[(+11)]
4.15am	0380 West Bromwich vs Manchester City	01 13.00	02 6.50	03 1.15	[M]	[⌚]	[(+11)]

Thu, 28 Jan 2021							
Time	Event	Home (1)	Draw (X)	Away (2)			
2.00am	0384 Burnley vs Aston Villa	01 3.30	02 3.30	03 1.92	[M]	[⌚]	[(+11)]
2.00am	0375 Chelsea vs Wolverhampton	01 1.53	02 3.80	03 5.00	[M]	[⌚]	[(+11)]
3.30am	0387 Brighton vs Fulham	01 2.00	02 3.10	03 3.30	[M]	[⌚]	[(+11)]
4.15am	0362 Everton vs Leicester	01 2.60	02 3.05	03 2.45	[M]	[⌚]	[(+11)]
4.15am	0377 Manchester Utd vs Sheffield Utd	01 1.23	02 5.20	03 10.00	[M]	[⌚]	[(+10)]

Fri, 29 Jan 2021							
Time	Event	Home (1)	Draw (X)	Away (2)			
4.00am	0374 Tottenham vs Liverpool	01 3.00	02 3.20	03 2.10	[M]	[⌚]	[(+11)]

PREDICTIONS:

Exact Score Prediction Using Home & Away Goals

- Similarly, we also tested the home total and away total goals of the matches.
- We also compare the odds created from the model and Singapore Pools'.

date	match	home	away	predicted home goals	0	1	2	3	4	5	6
2021-01-26	Newcastle_United_vs_Leeds_United	Newcastle United	Leeds United	2	0.146333	0.251213	0.496062	0.086157	0.019816	0.000418	2.142927e-07
2021-01-26	Crystal_Palace_vs_West_Ham_United	Crystal Palace	West Ham United	1	0.229635	0.408571	0.282619	0.071780	0.006631	0.000763	3.548526e-07
2021-01-26	West_Bromwich_Albia_vs_Manchester_City	West Bromwich Albion	Manchester City	1	0.274107	0.520543	0.176362	0.021963	0.006256	0.000767	2.692472e-07
2021-01-26	Southampton_vs_Arsenal	Southampton	Arsenal	1	0.125436	0.568376	0.121136	0.175364	0.008872	0.000816	3.261212e-07

date	match	home	away	predicted away goals	0	1	2	3	4	5	6
2021-01-26	Newcastle_United_vs_Leeds_United	Newcastle United	Leeds United	0	0.337836	0.273955	0.267260	0.108636	0.011249	0.000807	0.000257
2021-01-26	Crystal_Palace_vs_West_Ham_United	Crystal Palace	West Ham United	4	0.066043	0.051624	0.090841	0.057788	0.733177	0.000306	0.000221
2021-01-26	West_Bromwich_Albia_vs_Manchester_City	West Bromwich Albion	Manchester City	4	0.052679	0.091526	0.118555	0.275707	0.391460	0.056247	0.013825
2021-01-26	Southampton_vs_Arsenal	Southampton	Arsenal	2	0.125005	0.186355	0.327388	0.234166	0.113269	0.009975	0.003844

ENGLAND - PREMIER LEAGUE

JANUARY 27

FT	Crystal Palace	2 - 3	West Ham United	★
FT	Newcastle United	1 - 2	Leeds United	★
FT	Southampton	1 - 3	Arsenal	★
FT	West Bromwich Albion	0 - 5	Manchester City	★

PREDICTIONS:

Exact Score Prediction Using Home & Away Goals

- Here we compared a match's Home & Away Goals odds with Singapore Pools'.
- With this target variable, our prediction enables betting sites to produce three different betting method.

date	match	home	away	predicted home goals	predicted away goals							
					0	1	2	3	4	5	6	
2021-01-28	Tottenham_Hotspur_vs_Liverpool	Tottenham Hotspur	Liverpool	2.0	4.53	4.13	2.27	12.46	207.42	83.34	3707662.94	NaN
2021-01-28	Tottenham_Hotspur_vs_Liverpool	Tottenham Hotspur	Liverpool	NaN	3.89	3.33	4.04	5.29	426.23	1939.09	351.94	1.0

Fri, 29 Jan 2021

4.00am 0374 Tottenham vs Liverpool M Q B (+11)

09 | Total Goals Odd/Even

01 Odd	1.85	02 Even	1.80
----------	------	-----------	------

Fri, 29 Jan 2021

4.00am 0374 Tottenham vs Liverpool M Q B (+11)

05 | Total Goals

00 0 Goals	12.00	01 1 Goals	4.80	02 2 Goals	3.40
03 3 Goals	3.60	04 4 Goals	5.00	05 5 Goals	9.00
06 6 Goals	17.00	07 7 Goals	40.00	08 8 Goals	80.00

Fri, 29 Jan 2021

4.00am 0374 Tottenham vs Liverpool M Q B (+11)

04 | Pick The Score

Tottenham		Draw		Liverpool	
10 Tottenham 1-0	10.00	00 Draw 0-0	12.00	01 Liverpool 1-0	7.50
20 Tottenham 2-0	15.00	11 Draw 1-1	5.80	02 Liverpool 2-0	10.00
21 Tottenham 2-1	9.50	22 Draw 2-2	11.00	12 Liverpool 2-1	7.50

CONCLUSION

We are only able to get 53% accuracy and F1-Score for the Result prediction.

For Home and Away Goals predictions, we are only able to get 31.2% and 32.2%.

Football prediction is not as easy as it seems to be.

Other factors such match-fixing, players injuries, etc, can also affect the match outcome.

Moving Forward

More data like events during a match, players data and players to players data.

More targets can be explored.

Half-Time and Full-Time scores, which team to score first, which player to score first, number of corners, number of yellow cards etc.

THANK YOU

Do you have any questions?

And please do not use the predictions to gamble.

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by Freepik.

Please keep this slide for attribution.



SOURCES

- English Premier League – www.premierleague.com
- FBREF – www.fbref.com
- EA FIFA - <https://www.ea.com/games/fifa>
- Kaggle Dataset - <https://www.kaggle.com/stefanoleone992/fifa-21-complete-player-dataset>