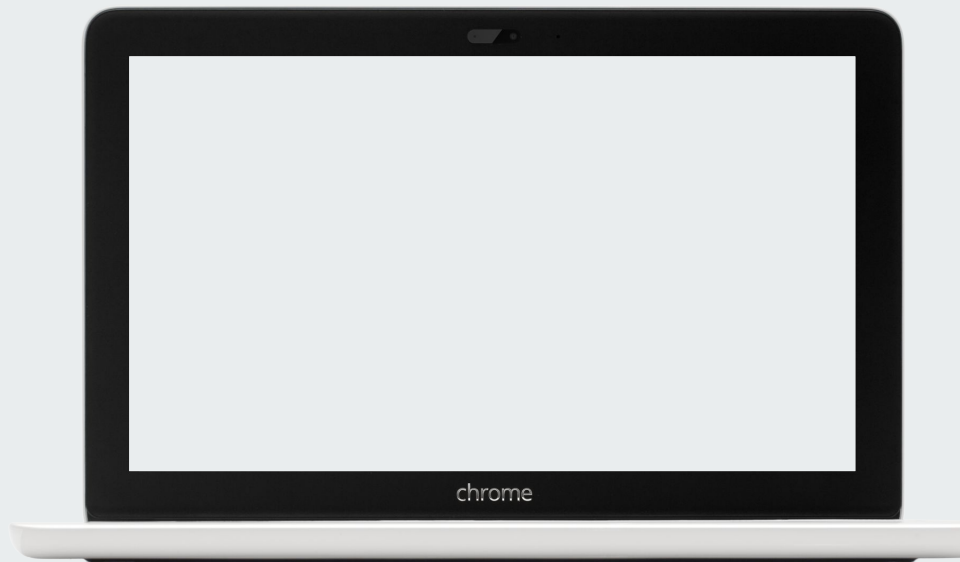




Predicting Goalscorers in Football

Modelling goalscorers to make profit in
the sports betting market



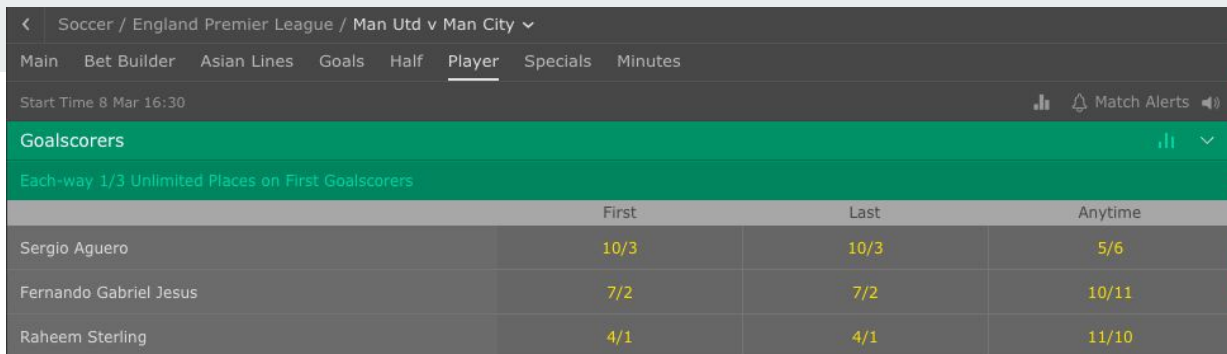


Problem statement

Quantifying a players chance of scoring in an individual football match .

By comparing my results to the betting market, I can quantify how successful I am and generate profit.

Example



Soccer / England Premier League / Man Utd v Man City

Main Bet Builder Asian Lines Goals Half Player Specials Minutes

Start Time 8 Mar 16:30

Match Alerts

Goalscorers

Each-way 1/3 Unlimited Places on First Goalscorers















	First	Last	Anytime
Sergio Aguero	10/3	10/3	5/6
Fernando Gabriel Jesus	7/2	7/2	10/11
Raheem Sterling	4/1	4/1	11/10

- The betting markets provides odds on a player to score
- Sergio Aguero is $\frac{5}{6}$ to score, this can be converted to a percentage. So we can say Sergio Aguero has a 54.64% chance to score.
- If my model predicts that Sergio Aguero will score 60% of the time. Then for that selection I can say I have a 9.8% edge on the market.
 - ◆ $0.6 / 0.5464 = 1.098$
- Therefore my goal is to build a model that highlights players where my probabilistic projections are higher than the betting markets odds.

Collecting data

01

- I scraped all the match commentary data from ESPN for over 20 leagues for the last 3 seasons.
- This totaled 642,045 shots taken by 42,533 players from 1,652 teams.

Chelsea	2	FT	1	Tottenham
81'		Foul by Giovani Lo Celso (Tottenham Hotspur).		
79'		Attempt saved. Tammy Abraham (Chelsea) left footed shot from very close range is saved in the centre of the goal.		
79'		Jorginho (Chelsea) wins a free kick on the left wing.		
79'		Foul by Giovani Lo Celso (Tottenham Hotspur).		
78'		Substitution, Tottenham Hotspur. Serge Aurier replaces Toby Alderweireld.		
78'		Substitution, Tottenham Hotspur. Dele Alli replaces Steven Bergwijn.		
78'		Hand ball by Giovani Lo Celso (Tottenham Hotspur).		
77'		Substitution, Chelsea. Willian replaces Ross Barkley.		
73'		Attempt blocked. Reece James (Chelsea) right footed shot from outside the box is blocked. Assisted by Ross Barkley.		
72'		Jorginho (Chelsea) wins a free kick in the defensive half.		
72'		Foul by Japhet Tanganga (Tottenham Hotspur).		
71'		Substitution, Chelsea. Tammy Abraham replaces Olivier Giroud.		
67'		Lucas Moura (Tottenham Hotspur) wins a free kick in the defensive half.		
67'		Foul by César Azpilicueta (Chelsea).		
66'		Attempt saved. Toby Alderweireld (Tottenham Hotspur) header from the centre of the box is saved in the top centre of the goal. Assisted by Harry Winks with a cross.		
66'		Corner, Tottenham Hotspur. Conceded by Marcos Alonso.		
66'		Attempt blocked. Erik Lamela (Tottenham Hotspur) left footed shot from outside the box is blocked.		

Creating Metrics

02

- The data gives me various shot locations, from this I can see the success rate of scoring a goal from each location
- I can also look at each individual players goal percentage from each location
- From this data I could calculate a expected goals metric for each player

	shot_location	shots	goal%
0	header - centre of the box	35593	0.11
1	header - difficult angle	571	0.03
2	header - side of six yard box	4360	0.19
3	header - side of the box	1701	0.03
4	header - very close range	4196	0.45
5	shot - centre of the box	53557	0.23
6	shot - difficult angle	4841	0.10
7	shot - difficult angle and long range	1336	0.08
8	shot - long range	3566	0.03
9	shot - outside the box	99134	0.04
10	shot - side of the box	40612	0.09
11	shot - side of the six yard box	5452	0.29
12	shot - very close range	6145	0.66

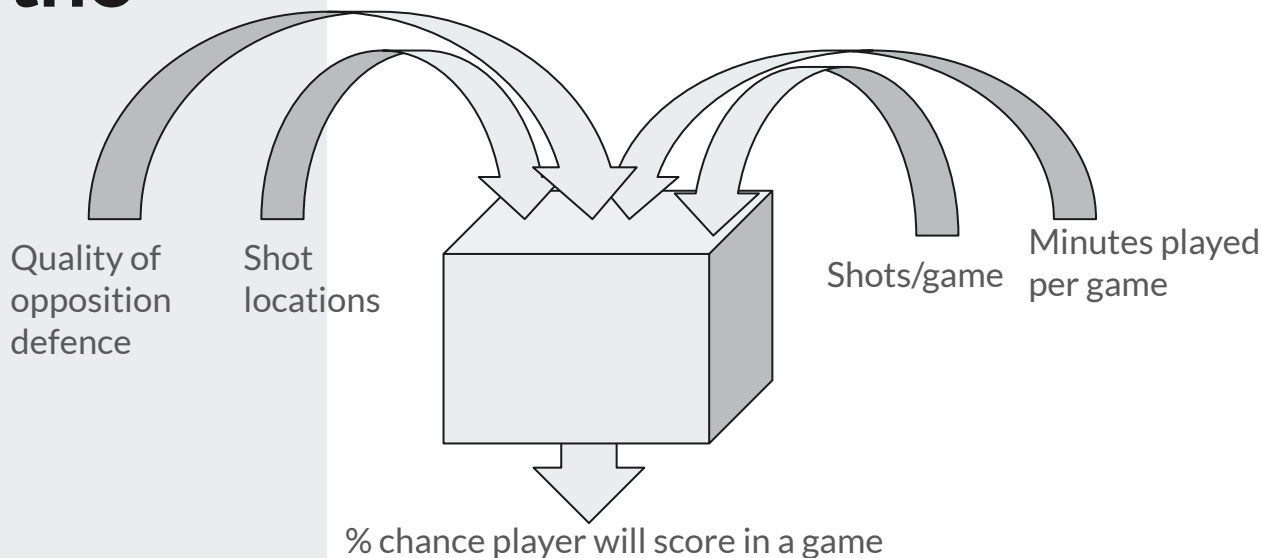
shot_location	player	shots	goal%	shot_location	player	shots	goal%
header - centre of the box	Lionel Messi	17	0.00	header - centre of the box	Cristiano Ronaldo	64	0.09
header - very close range	Lionel Messi	1	0.00	header - side of six yard box	Cristiano Ronaldo	6	0.17
shot - centre of the box	Lionel Messi	117	0.35	header - side of the box	Cristiano Ronaldo	2	0.00
shot - difficult angle	Lionel Messi	12	0.17	header - very close range	Cristiano Ronaldo	7	0.57
shot - difficult angle and long range	Lionel Messi	1	0.00	shot - centre of the box	Cristiano Ronaldo	102	0.22
shot - long range	Lionel Messi	2	0.00	shot - difficult angle	Cristiano Ronaldo	9	0.00
shot - outside the box	Lionel Messi	205	0.04	shot - difficult angle and long range	Cristiano Ronaldo	7	0.00
shot - side of the box	Lionel Messi	44	0.07	shot - long range	Cristiano Ronaldo	4	0.00
shot - side of the six yard box	Lionel Messi	10	0.20	shot - outside the box	Cristiano Ronaldo	128	0.01
shot - very close range	Lionel Messi	8	0.88	shot - side of the box	Cristiano Ronaldo	91	0.10
				shot - side of the six yard box	Cristiano Ronaldo	10	0.30
				shot - very close range	Cristiano Ronaldo	16	0.44

player	xg	goals	dif
Robert Lewandowski	69.444	63	6.444
Lionel Messi	64.354	64	0.354
Mohamed Salah	61.213	62	-0.787
Cristiano Ronaldo	57.458	53	4.458
Kyllian Mbappé	56.841	54	2.841
Paulinho	52.016	48	4.016
Ciro Immobile	51.274	50	1.274
Pierre-Emerick Aubameyang	50.553	52	-1.447
Sergio Agüero	50.046	50	0.046
Harry Kane	48.636	50	-1.364
Luis Suárez	48.092	53	-4.908
Lawrence Shankland	47.679	63	-15.321
Sadio Mané	46.459	45	1.459
Timo Werner	45.805	45	0.805
Jarrod Bowen	44.742	47	-2.258
Aleksandar Mitrovic	43.985	44	-0.015
Romelu Lukaku	43.464	41	2.464
Edinson Cavani	43.139	41	2.139
Ollie Watkins	41.986	43	-1.014
Mauro Icardi	41.750	41	0.750

Modelling and Automating the Data

03

- I then used various statistical techniques to model this data (linear regression, logistic regression, conditional algorithms, monte carlo simulations)
- I also wanted to make the model fully automated so I could click a button and it would give me the suggested bets.





Outputs

04

- Merge the odds dataset with the goalscorer dataset
- Calculate the selections which are projected to be profitable
- Output the selections that are profitable

date	league	player	H	A	Odds	Games	Avg/mins	Proj_ROI
2020-01-11 14:00:00	2019-20 Italian Serie A	Radja Nainggolan/Cagliari	Cagliari	AC Milan	6.50	15	80.67	46.25%
2020-01-11 15:00:00	2019-20 English League Two	Joss Labadie/Newport County	Scunthorpe	Newport County	8.50	17	79.24	126.95%
2020-01-11 15:00:00	2019-20 English League One	Ian Henderson/Rochdale	Rochdale	Bolton	2.60	18	83.72	28.7%
2020-01-11 15:00:00	2019-20 English League One	Ivan Toney/Peterborough United	Peterborough United	Gillingham	2.70	20	88.75	51.2%
2020-01-11 15:00:00	2019-20 English League Championship	Jake Cooper/Millwall	Stoke	Millwall	17.00	24	90.00	97.2%
2020-01-11 15:00:00	2019-20 English League Two	Padraig Amond/Newport County	Scunthorpe	Newport County	3.30	21	71.10	49.16%
2020-01-11 15:00:00	2019-20 English League Two	Josh Gordon/Walsall	Cheltenham	Walsall	4.00	22	65.14	38.8%
2020-01-11 15:00:00	2019-20 English League Championship	Alex Mowatt/Barnsley	Barnsley	Huddersfield	8.00	23	88.13	72.8%



Results!

05

→ After optimising and subsetting the dataset I have the following results.

Bets: 152
Profit: 45.15
Return on Investment: 29.7%



Picks for the weekend

06

Picks here.....

Questions?