"The Leicester City Fairytale?": Utilizing New Soccer Analytics Tools to Compare Performance in the 15/16 & 16/17 EPL Seasons

Hector Ruiz
STATS
Barcelona, Spain
hruiz@stats.com

Paul Power
STATS
Leeds, United Kingdom
paul.power@stats.com

Xinyu Wei
STATS
Sydney, Australia
xwei@stats.com

Patrick Lucey
STATS
Chicago, USA
plucey@stats.com

ABSTRACT

The last two years have been somewhat of a rollercoaster for English Premier League (EPL) team Leicester City. In the 2015/16 season, against all odds and logic, they won the league to much fan-fare. Fast-forward nine months later, and they are battling relegation. What could describe this fluctuating form? As soccer is a very complex and strategic game, common statistics (e.g., passes, shots, possession) do not really tell the full story on how a team succeeds and fails. However, using machine learning tools and a plethora of data, it is now possible to obtain some insights into how a team performs. To showcase the utility of these new tools (i.e., expected goal value, expected save value, strategy-plots and passing quality measures), we first analyze the EPL 2015/16 season which a specific emphasis on the champions Leicester City, and then compare it to the current one. Finally, we show how these features can be used to predict future performance.

KEYWORDS

Prediction, Sports Analytics, Unstructured Data

ACM Reference format:
Hector Ruiz, Paul Power, Xinyu Wei, and Patrick Lucey. 2017. "The Leicester City Fairytale?": Utilizing New Soccer Analytics Tools to Compare Performance in the 15/16 & 16/17 EPL Seasons. In *Proceedings of KDD'17, August 13–17, 2017, Halifax, NS, Canada., ,* 10 pages.
DOI: 10.1145/3097983.3098121

1 INTRODUCTION

The fairytale of Leicester City winning the English Premier League (EPL) in 2015/16 has been well documented. From starting the season with having odds of 5000-1; uncovering hidden gems of N'Golo Kanté and Riyad Mahrez; the goal-scoring form of Jamie Vardy; to the minimal number of injuries as well as relatively small number of games they had to play compared to their competitors - there were many storylines and explanations that resonated with the footballing community.

In terms of playing style or strategy, there has been also much made of their direct or counter attacking nature [1, 2, 7]. Compared to the previous champions of the last five years (see Table 1), Leicester City tended to cede possession (43% vs 55%) and hit teams on the counter attack. However, when you look at the *expected goal value* (xG)[1] [9] in terms of attacking prowess, they were not superior to the other top teams. With respect to the chances they created, they were expected to score approximately 66 goals which was close to the actual 68 goals they scored, which was about average across the league. However, when you compare their goal-scoring effectiveness to the previous champions of the past 5 years, we can see the differential of +2 is minuscule when compared to the Manchester City and United teams who had a differential ranging from +11 to +23 goals. This result suggests that even though Vardy and Mahrez had stellar seasons, in terms of goal-scoring execution, they were far behind Sergio Aguero (11/12 & 13/14), and Robin Van Persie (12/13) in the seasons when their team won the premiership.[2] Even in the 2014-2015 season, Chelsea had a differential of nearly +10 goals which correlated with the outstanding seasons of Diego Costa and Eden Hazard.

So if Leicester City were not more effective than other teams offensively due to their counter-attacking/direct style of play, how did they win the 2015/16 title? Well there are two reasons and both can be explained by the expected goal differential (one on the offensive side, and the other on the defensive side). The first reason: **the absence of the normal contenders**. As mentioned above, normally the team who wins the league catches fire and has a player or two who is significantly better than the other players in the league. Even though Leicester did not have those players, neither did any of the other teams (Tottenham were the closest team with a +7 differential).

The second reason is that **Leicester had the most effective defence**. Not only were they the most effective for that season, they were by far the best across the last five seasons. Even though they conceded 36 goals, the expected goal value model suggested

[1]Expected goal value (xG) refers to the likelihood of "the average player" scoring from a given situation.
[2]Expected goal value can be used to rate the effectiveness/talent of a striker (i.e., in a given situation a player like Aguero or Van Persie may be more likely to score compared to the average EPL player).

Table 1: Key statistics showing the offensive and defensive effectiveness of the top EPL teams across the last 5 seasons. The "*" symbol next to GF and GA denotes that own goals do not count towards the final goal tally.

Season	Rank	Team	Points	Poss %	GF*	xGF	GF*-xGF	GA*	xGA	GA*-xGA
2011/12	**1**	**Manchester City**	**89**	**55.7**	**91**	**81.5**	**9.5**	**28**	**31.0**	**-3.0**
	2	Manchester United	89	55.8	87	75.4	11.6	31	36.2	-5.2
	3	Arsenal	70	57.6	74	72.9	1.1	44	37.8	6.2
	4	Tottenham Hotspur	69	55.6	66	66.1	-0.1	40	43.2	-3.2
2012/13	**1**	**Manchester United**	**89**	**54.2**	**80**	**68.9**	**11.1**	**39**	**35.8**	**3.2**
	2	Manchester City	78	56.0	64	70.1	-6.1	32	31.7	0.3
	3	Chelsea	75	53.2	71	62.3	8.7	37	38.9	-1.9
	4	Arsenal	73	55.2	69	64.2	4.8	37	38.1	-1.1
2013/14	**1**	**Manchester City**	**86**	**56.0**	**98**	**78.8**	**19.2**	**35**	**31.7**	**3.3**
	2	Liverpool	84	53.9	97	75.0	22.0	45	40.5	4.5
	3	Chelsea	82	55.0	69	68.3	0.7	26	33.1	-7.1
	4	Arsenal	79	54.0	68	54.7	13.3	39	41.0	-2.0
2014/15	**1**	**Chelsea**	**87**	**54.7**	**72**	**63.5**	**8.5**	**31**	**35.9**	**-4.9**
	2	Manchester City	79	56.9	82	77.6	4.4	35	39.1	-4.1
	3	Arsenal	75	54.7	69	65.1	3.9	35	34.4	0.6
	4	Manchester United	70	60.3	60	51.1	8.9	36	39.0	-3.0
2015/16	**1**	**Leicester City**	**81**	**43.4**	**68**	**66.1**	**1.9**	**36**	**46.7**	**-10.7**
	2	Arsenal	71	54.8	62	69.1	-7.1	34	35.5	-1.5
	3	Tottenham Hotspur	70	56.7	68	63.0	5.0	32	35.1	-3.1
	4	Manchester City	66	55.3	71	67.0	4.0	39	37.0	2.0

that they should have conceded 46.5 goals, a differential of over -10.7 goals (far right of Table 1).

To give context of how unprecedented both these phenomena were, we plotted a histogram of both the offensive and defensive expected goal differential aggregated for all English Premier League teams across the last 5 seasons which we show in Figure 1. In Figure 1(a), we show a histogram of the offensive effectiveness of all the teams (N=100) across the last 5 seasons (we have highlighted where the champions of that season were, the red bars highlight the mean, one-standard deviation and two standard deviations). As can be seen, the season that Manchester City had in 2013/14, they were well outside two standard deviations (the other team was Liverpool that year with Suarez and Sturridge), while Leicester was hovering around the mean. When we look at the defensive effectiveness however (Figure 1(b)), Leicester was essentially the Manchester City counter-part, but this time on the defensive side. Which begs the question, what can explain Leicester City defensive effectiveness? There are a host of possible reasons:

(1) **Goal-keeping**: Similar to the Manchester City offensive example where the strikers were far better than the average striker, the measure could be down to exceptional saves that Kasper Schmeichel made compared to the other keepers in the league (Section 2).
(2) **Defensive Strategy**: As been noted previously, the average expected goal value varies depending on the game-state. For example, the likelihood of scoring from a counter-attack is much higher than normal build up. Leicester City may have been far more effective defending a type of play and invited teams to attack a certain way which suited their defensive strategy. We show how we can understand this via *"Strategy Plots"* (Section 3).
(3) **Passing Disruption**: In 50-50 situations, Leicester may have been able to stop plays before they started. We do this analysis by measuring the quality of each pass (Section 4).

In this paper, we will explore each reason using advanced tools and metrics which can help explain this phenomenon. Additionally, we use these same tools to explain why Leicester City are not achieving the same heights as they did last year (Section 5). Additionally, we will show how we can *"recommend"* future performances based on team strategy via our *"recommendation engine"* (Section 6).

2 MEASURING GOAL-KEEPING PERFORMANCE

Expected goal value measures the quality of a chance or, given a situation, it allows us to estimate the likelihood of the average player in the league to score a goal. However, it does not describe the execution of the shot. Without the fine-grain information about the shot trajectory at the frame-level (i.e., x,y,z), as well as the goal-keeping location and motion, determining goal-keeping effectiveness is challenging. Despite that, one piece of information which we do have is the "shot-on-target" and "shot-off-target" label. To see if Kasper Schmeichel was the reason why Leicester City were significantly more effective than any other team in recent EPL history, we can see if his save-rate (i.e., number of shots-on-target/goals) was higher than other goal-keepers.

Table 2: Leicester City vs EPL average for shots against, split into goals conceded (G) and missed (M). The latter are broken down into blocked, saved and off target.

Team	Shots	G	M	Blocked		Saved		Off target	
EPL Average	501.5	49.8	451.7	176.6	39.1%	111.4	24.7%	163.8	36.3%
Leicester City	532	36	496	225	45.4%	101	20.4%	170	34.3%
* **Leicester City**	532	36	496	194	39.1%	122	24.6%	180	36.3%
				$\boxed{\mathbf{31^*}}$		$\boxed{-21}$		$\boxed{-10}$	

***p-value: 0.0025**

Table 2 breaks down the number of missed shots by their possible outcomes (block, save or shot-off-target), and compares the

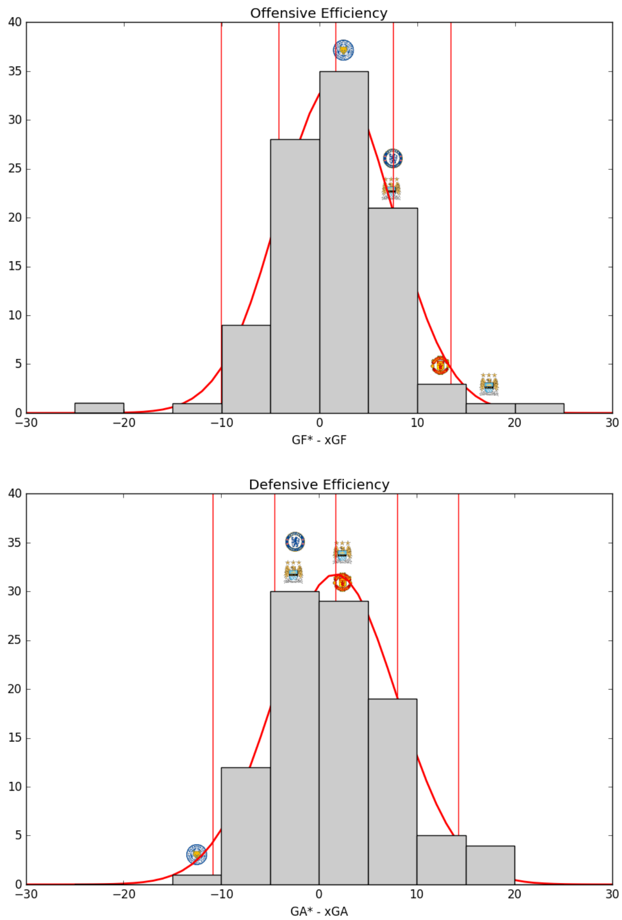

Figure 1: (a) Shows the goals-for/expected-goals-for (GF*-xGF) differential for all EPL teams across the last 5 seasons (N=100). The middle red-line corresponds to the mean, the red-lines either side of the middle are one standard deviation, and the widest red-lines corresponds to two standard-deviations. Manchester City and Liverpool from the 13/14 season were outside two standard-deviations. (b) Shows the goals-against/expected-goals-against (GA*-xGA) — Leicester City for the 15/16 season were two standard deviations outside the mean.

proportion of shots missed by Leicester's opposition in each of those categories with the average league values. We also propose a hypothetical comparison: the third row in the table contains the figures that Leicester would have had assuming that the proportion of missed shots that were blocked, saved and off target were the same as the league average ones. Such construct would have resulted in 31 fewer shots blocked, 21 more shots saved and 10 more shots off target, which undoubtedly emphasizes the defensive efforts of Leicester's field players, who went above and beyond to negate the scoring chances of their opponents.

Even though the Leicester goalkeeper had fewer shots to save than other keepers in the league, we want to check if the shots he did save were high-quality or not. Or in another way, we want to ask the question: "given a shot on target, what is the likelihood that the average goal-keeper will save it?"

Table 3: The ranking of the 20 EPL teams for the 2015/16 season in terms of goal-keeping (N=number of shots-on-target (Conceded + Saved), xS=expected number of saves).

Rank	Team	N	Conceded	Saved	xS	Saved-xS
1	**Watford**	181	48	133	125.0	8.0
2	**Leicester City**	137	36	101	96.4	4.6
3	**Arsenal**	157	34	123	118.6	4.4
4	**West Ham United**	162	49	113	109.3	3.7
5	**Manchester United**	120	32	88	84.9	3.1
6	**Swansea City**	168	49	119	116.1	2.9
7	**Sunderland**	214	61	153	150.7	2.3
8	**Crystal Palace**	164	48	116	114.5	1.5
9	**Southampton**	141	41	100	99.2	0.8
10	**Tottenham Hotspur**	124	32	92	91.3	0.7
11	**Manchester City**	124	39	85	84.9	0.1
12	**Everton**	185	55	130	130.0	0.0
13	**Stoke City**	176	55	121	121.7	-0.7
14	**Chelsea**	180	52	128	128.9	-0.9
15	**West Bromwich Albion**	156	46	110	112.7	-2.7
16	**Norwich City**	177	66	111	116.0	-5.0
17	**Newcastle United**	180	63	117	123.4	-6.4
18	**Liverpool**	135	49	86	92.8	-6.8
19	**Aston Villa**	187	73	114	123.7	-9.7
20	**Bournemouth**	154	67	87	97.9	-10.9

2.1 Expected Save Value (xS)

Expected Save Value (xS) measures the likelihood of a goal-keeper making a save from a shot [4, 10]. It may seem that this would be just the compliment of expected goal value (i.e., xS = 1 - xG), however there is a slight difference. In training our classifier for estimating the xS, our data samples consist of only the shots-on-target (for xG, the training samples consist of both shots-on-target and shots-off-target). Similar to our xG model, we trained a neural network on the event-based features to estimate the probability of a goal-keeper making a save (see section 3 for more details on the dataset and features used). In Table 3, we show the list of the goal-keepers of all teams, ranked according to their differential between the number of saves they made (Saved) and the expected saves (xS). In the last column, we can see that the Watford keeper (Heurelho Gomes) was the most effective keeper in the league (+8.0), followed by Leicester's Kasper Schmeichel (+4.6), then the combination Petr Cech/David Ospina for Arsenal (+4.4).

From this analysis, we can see that Schmeichel played a key role, with his performance accounting for 4.6 of the 10.7 goal differential for Leicester. What about the remaining 6.1 goals? In addition to the number of blocks they had, it is also about the strategy they deployed.

3 MEASURING STRATEGY EFFECTIVENESS DIRECTLY FROM DATA

To measure the effectiveness of a strategy, we use the expected goal value (xG) model as our primary measurement tool. Our goal expectancy models estimate the conversion probability of a shot using the following set of features to describe it:

- **X, Y, X^2, Y^2, XY**: Shot location coordinates, normalized to take into account the direction of attack of the shooting team.

- **Goal angle and distance**: Angle and distance from the shot location to the center of the goal.
- **Previous angle and distance**: Angle and distance from the previous ball touch to the shot location.
- **Open-play footed, headed, free kick, penalty**: Four mutually-exclusive binary features to indicate the type of shot.
- **Cutback**: Binary feature to indicate whether the shot is preceded by a cutback pass.
- **First touch**: Binary feature to indicate whether the shot was a first-touch ball contact.

The predictions are the result of averaging an ensemble of three classifiers (logistic regression, random forest and multilayer perceptron) built using a dataset comprising all shots taken during the five most recent EPL seasons (2011/12 through 2015/2016, n=51,388). Table 4 shows the distribution of the shot sample used for our expectancy models by season and outcome.

Accurate expected goal value estimation provides valuable information about teams' performances. Table 1 and Figure 1, for example, give a very clear message that highlights the uniqueness of the 2015/16 EPL: unlike in previous occasions, where the eventual champions unequivocally excelled in their offensive performance, the analysis of expected versus actual goals scored suggests that this time the defining factor that gave Leicester the title was defensive effectiveness.

An abundance of work has focused on using xG for offensive analysis [5, 9]. Similarly, it can be used for defensive analysis. The usefulness of xG is that it allows analysts to obtain a relative picture on how teams attacked and defended compared to the rest of the league. However, as soccer is a very strategic game consisting of many different ways to attack/defend, we thought it would be prudent to describe a dictionary of "methods of scoring" or "shot-playbook". As such categories are very nuanced and require a high-level of domain specific knowledge, we first worked with soccer experts who helped us create a dictionary which consists of 14 elements, D = Build Up, Build Up*, Corner Kick, Counter Attack, Counter Attack*, Cross, Direct Play, Direct Play*, Free Kick Shot, From Cross, From Free Kick, Other, Penalty Shot, Throw In — depictions of the shot-dictionary are shown in Appendix A. To apply the shot dictionary to all shots across the 5 seasons of data we had, we first asked the domain experts to label a subset of exemplars for each shot category by watching video. We synchronized the event data with the video, and using the features and classifiers for our xG and xS estimates we learnt classifiers for each class label. It should be noted that "*" in certain shot types means that they have been assigned from being ambiguous samples to that style during the active learning process.

Table 4: Breakdown of the shot dataset used for training our expected goal model. The meaning of the headings is the same than in Table 2. For our expected save model, the shot sample is restricted to goals (G) and Saved (n=16,721).

Season	Shots	G	M	Blocked	Saved	Off target
2011/12	10671	1023	9648	3456	2427	3765
2012/13	10372	1014	9358	3212	2453	3693
2013/14	10319	1008	9311	3269	2410	3632
2014/15	9997	946	9051	3520	2218	3313
2015/16	10029	995	9034	3531	2227	3276
Total	51388	4986	46402	16988	11735	17679

Table 5: Leicester City vs EPL average for number of shots per 90 minutes and average xG for the different shot-types they conceded (i.e., defensively).

Shot Type	EPL Average		Leicester City	
	Shots/90	Avg. xG	Shots/90	Avg. xG
Build Up	0.982	0.069	0.835	0.072
Build Up*	0.963	0.074	1.080	0.061
Corner Kick	1.659	0.091	2.161	0.092
Counter Attack	0.697	0.135	0.663	0.129
Counter Attack*	0.687	0.070	0.589	0.060
Cross	1.136	0.157	1.277	0.127
From Cross	1.194	0.094	1.621	0.091
Direct Play	0.766	0.101	0.516	0.127
Direct Play*	0.204	0.061	0.123	0.045
Free Kick Shot	0.487	0.075	0.344	0.076
From Free Kick	0.768	0.094	0.663	0.054
Other	2.266	0.073	2.578	0.063
Penalty Shot	0.112	0.777	0.098	0.775
Throw In	0.409	0.061	0.516	0.037
Total	**12.330**	**0.097**	**13.064**	**0.088**

We then applied each classifier to each example and obtained initial results which we evaluated. The initial performance was poor, but using active learning [6] we were eventually able to obtain close to perfect assignment of shots to each one of the 14 classes. The intuition behind active learning is that we get the classifiers to first only label the easy examples, and then use a human expert to label the difficult/ambiguous examples. The benefit of this approach is that once the human intervenes, we can retrain the classifiers which in essence makes the classifier better and more robust after each iteration.

Due to the different style and strategy of teams, using our shot-dictionary, we were able to decouple each team's behaviour into each of the 14 categories. In Table 5, we show how Leicester City defended across the 2015/16 season compared to the league average across two dimensions: i) number of shots, and ii) the average expected goal value.

3.1 Introducing "Strategy Plots" to Compare Strategic Performance

In terms of analyzing team behavior, there are two attributes which are important: i) number of shots, and ii) shot effectiveness. As evidenced in this paper so far, tables (e.g., Tables 1-5) can disseminate information adequately but when the number of variables increases doing meaningful analysis can be tough. Our "strategy plot" is based on a "Hinton Diagram" [9], named after famous machine learning researcher Geoffrey Hinton who used this technique to visualize hundreds or thousands of connections, as viewing them within tables was meaningless. Even though we only have 14 attributes, the basic premise is the same, since viewing 14 attributes across 20 teams (280 values) can be equally meaningless. For Leicester City's values in Table 5, we show the visualization of such values via a strategy plot in Figure 3.

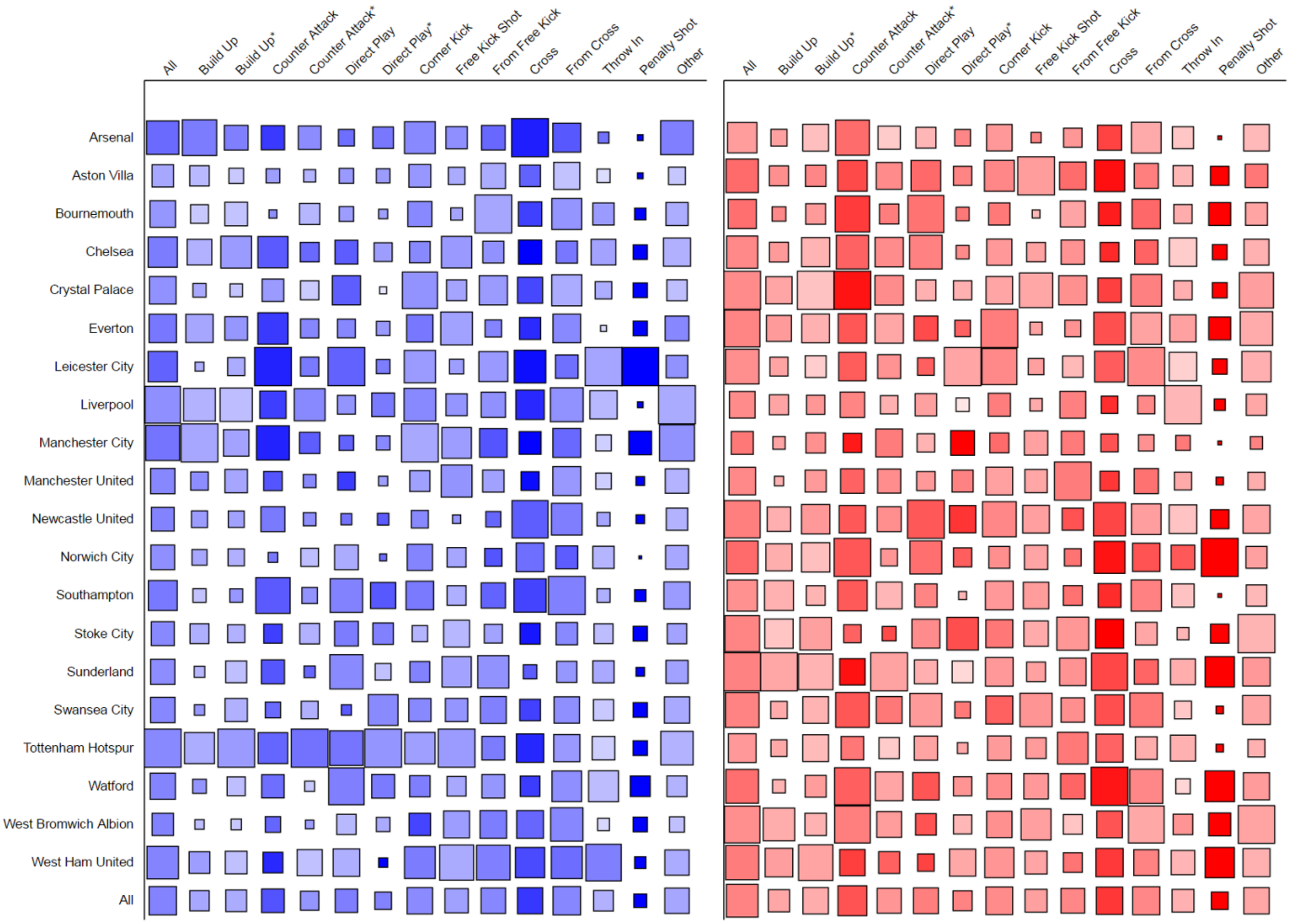

Figure 2: Strategy plot of the 2015/16 EPL season comparing all 20 teams offensive (left) and defensive (right) effectiveness for all 14 shot types. The size of the square relates to how many shots were made/conceded for each shot type. The intensity of the color corresponds to the effectiveness with respect to xG (i.e., light = low xG, dark = high xG).

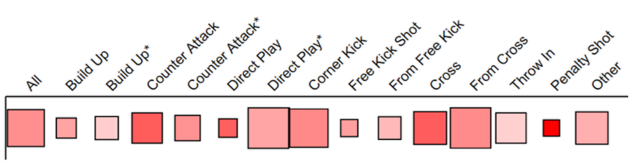

Figure 3: Strategy plot showing the number of shots that Leicester conceded as well as how effective they were in this context. The size is normalized to the maximum value for that shot category.

The interpretation of a strategy plot is quite simple — the size of the square corresponds to number of shots (relative to the maximum value observed in the league for that category) and intensity of color corresponds to the effectiveness of each method with respect to xG (light-red=low-xG, dark-red=high-xG). So in terms of defending, Leicester conceded a lot of shots via direct-play*, corner kicks and from crosses but were very effective in dealing with them (light-red). The effectiveness for the contexts where they gave up most shots can account for the additional 6 goals they were expected to concede (see Table 6 for more analysis).

Strategy plots also can be used to quickly glean strategic effectiveness at the league-level. In terms of defence we can see in Figure 2, that Liverpool and Manchester City gave up the least number of shots and these were evenly distributed (after normalization) across the 14 types. In terms of effectiveness, Manchester City were vulnerable to counter-attacks and direct-play*, while Liverpool were not effective in defending crosses. It can also be seen that Tottenham gave up more shots, but were pretty effective in all defending situations (light-red), which emphasizes their defensive strength in the 2015/16 season. We can also see that while Crystal Palace gave up a lot of shots, they were least effective at defending counter-attacks. Similarly, we can see that Watford gave up a lot of shots from crosses and did not defend them well as the xG against them was relatively high compared to other teams.

The left-hand side of Figure 2 shows the strategy plots for the offensive performance of all 20 teams. What immediately jumps out is that Leicester City had a lot of counter-attacks and were very effective using this method of scoring. They were also the team with the most penalties, 13 throughout the season. We can also see that Tottenham created many chances in many different ways, and were particularly strong from crosses. We can also see that Arsenal, Liverpool, Manchester City created a lot a chances in many different ways too. Manchester United on the other-hand were rather offensively limp, with not many chances being created – similar to other offensively deficient teams like, Stoke, Norwich, Newcastle and Aston Villa.

4 MEASURING TEAM DEFENSE VIA PASSING ANALYSIS

The defensive acumen of a team goes beyond how they defend in scoring situations. It is not just the number of tackles, regains or interceptions made by a team, either. Xabi Alonso was once quoted as saying, "Tackling is not really a quality, it's more something you are forced to resort to when you don't have the ball". Therefore, to measure defending we have to be able to capture what the defence is "forcing" the team in possession to do. With current machine learning techniques, we can measure the difficulty of a pass – which can give us an indication on whether a defence is "forcing" an attacker to attempt a more difficult pass than necessary or turning low probability transition situations into high probability transition situations such as a 50-50 challenge.

4.1 Passing Difficulty Model

To measure such things, we have created a pass difficulty model which assigns a probability to the completion of any pass based on the context of the situation such as: where the ball is on the pitch, where the pass is intended to go and how much pressure the receiver and passer are under. Our model has been trained on over 480k examples with labels pass-made=1, pass-not-made=0, and is similar to those implemented in [3, 8, 11]. Having a pass-quality model allows each pass to be rated by its difficulty to complete and when aggregated provides a robust metric to capture both attacking and defending ability at a team and player level. Similar to our strategy plots, we can visualize how teams compare in the relative number of passes conceded for different difficulties. In Figure 4, we examine the proportion of passes intercepted depending on passing difficulty. We have broken these into 5 categories (1-20%, 21-40%, 41-60%, 61-80% and 81-100%), with 1% being the most difficult pass and 100% being the easiest. The intensity represents how often a team has regained the ball in the bin compared to the top team. Darker shows higher regain percentage (normalized to the maximum in the category), and vice versa. As such, all teams are about the same in bin 80-100% but there is a marked difference in all other bins between the top and bottom teams.

From Figure 4, we also immediately see that Leicester City exceled (ranking 1st) at regaining the ball for passes in the 21-40%, 41-60% and 61-80% bands. At an individual level, N'Golo Kanté was the highlight of the team; he was the midfiel player with the highest interception rate for passes in the <80% bands. This insight fits with the media narrative of Leicester City being highly compact in defence with Kanté being the key in breaking up attacks in the middle of the pitch.

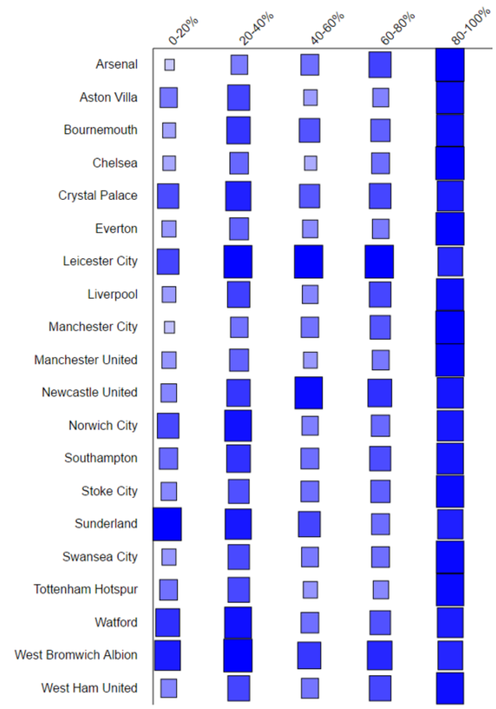

Figure 4: Passing diagram showing interception rate for each pass difficulty band. Size of square is the number of passes observed and intensity of color represents disruption percentage.

5 ANALYSIS OF THE CURRENT 2016/17 SEASON

After achieving the unthinkable last season, it was not clear what to expect from Leicester City a year on. There was a generally accepted opinion that the team would not be able to repeat their great feat, and that having to balance the domestic competition with the Champions League would see them regress to places nearer the middle of the table, rather than the top. What not many expected, however, was that they would struggle so much; at the moment of writing this paper, Leicester are 17th in the table, just one point above the relegation zone.

One question immediately comes to mind: *what has changed that explains this drop in performance?* Table 7 shows the current[3] league standings. Comparing these values with those in Table 1, there is one striking fact that jumps to the eye: the difference between the number of expected and actual goals conceded has gone from -10.7 (-0.28/game) to +8.5 (+0.34/game). What this means in practice is that, even though there is not a large difference in the volume of expected goals conceded (1.34/game compared to 1.23/game last season), their opponents are now being able to materialize those

[3]The analysis of the 2016/17 season is current as of February 15, 2017, after the first 25 rounds of games

Table 6: Comparison of Leicester's current and past seasons in terms of the type of shots they conceded.

Shot Type	2015/16 Leicester City				2016/17 Leicester City			
	Shots/90m	Avg. xG	Goals/90m	eff	Shots/90m	Avg. xG	Goals/90m	eff
Build Up	0.835	0.072	0.074	0.088	1.049	0.084	0.075	0.071
Build Up*	1.080	0.061	0.000	0.000	1.049	0.056	0.037	0.036
Corner Kick	2.161	0.092	0.172	**0.080**	1.536	0.110	0.225	**0.146**
Counter Attack	0.663	0.129	0.025	**0.037**	0.412	0.170	0.112	**0.273**
Counter Attack*	0.589	0.060	0.049	**0.083**	0.787	0.088	0.112	**0.143**
Cross	1.277	0.127	0.123	0.096	1.236	0.119	0.150	0.121
From Cross	1.621	0.091	0.123	0.076	2.210	0.084	0.150	0.068
Direct Play	0.516	0.127	0.049	**0.095**	1.161	0.111	0.262	**0.226**
Direct Play*	0.123	0.045	0.000	0.000	0.187	0.026	0.000	0.000
Free Kick Shot	0.344	0.076	0.025	0.071	0.525	0.058	0.037	0.071
From Free Kick	0.663	0.054	0.025	**0.037**	0.674	0.096	0.150	**0.222**
Other	2.578	0.063	0.098	0.038	2.285	0.060	0.150	0.066
Penalty Shot	0.098	0.775	0.098	1.000	0.112	0.771	0.112	1.000
Throw In	0.516	0.037	0.025	0.048	0.450	0.030	0.000	0.000
Total	**13.064**	**0.088**	**0.884**	**0.068**	**13.675**	**0.092**	**1.574**	**0.115**

Table 7: 2016/17 EPL table up to (and including) week 25.

Rank	Team	Points	GF*	xGF	GF*-xGF	GA*	xGA	GA*-xGA
1	Chelsea	60	51	36	15	17	17.8	-0.8
2	Manchester City	52	47	49	-2	28	20.1	7.9
3	Tottenham Hotspur	50	44	42	2	17	25.4	-8.4
4	Arsenal	50	51	50	1	27	29.6	-2.6
5	Liverpool	49	53	49.9	3.1	30	24.8	5.2
6	Manchester United	48	38	42.2	-4.2	20	20.1	-0.1
7	Everton	41	37	35.5	1.5	26	32.3	-6.3
8	West Bromwich Albion	37	34	28.4	5.6	30	36.4	-6.4
9	Stoke City	32	27	29.9	-2.9	32	32.4	-0.4
10	West Ham United	32	33	33.4	-0.4	43	41.1	1.9
11	Southampton	30	26	36.6	-10.6	31	29.5	1.5
12	Burnley	30	27	23.8	3.2	35	38.9	-3.9
13	Watford	30	28	24.8	3.2	41	38.6	2.4
14	Bournemouth	26	35	30.2	4.8	46	41	5
15	Swansea City	24	31	30	1	50	46.4	3.6
16	Middlesbrough	22	18	21.2	-3.2	27	36.2	-9.2
17	Leicester City	21	23	28.2	-5.2	42	33.5	8.5
18	Hull City	20	21	21.9	-0.9	46	52.7	-6.7
19	Crystal Palace	19	31	34	-3	46	32.6	13.4
20	Sunderland	19	23	24.5	-1.5	44	42.1	1.9

Table 8: The ranking of the 20 EPL teams for the 2016/17 season in terms of goal-keeping (N=number of shots-on-target (Conceded + Saved), xS=expected number of saves).

Rank	Team	N	Conceded	Saved	xS	Saved-xS
1	Burnley	145	35	110	101.8	8.2
2	Tottenham Hotspur	73	17	56	47.8	8.2
3	West Bromwich Albion	112	30	82	74.7	7.3
4	Hull City	146	46	100	93.4	6.6
5	Everton	97	26	71	64.4	6.6
6	Middlesbrough	104	27	77	72.1	4.9
7	Arsenal	100	27	73	68.6	4.4
8	Manchester United	81	20	61	56.7	4.3
9	Chelsea	72	17	55	53.1	1.9
10	Watford	122	41	81	80.1	0.9
11	Sunderland	160	44	116	115.1	0.9
12	Liverpool	77	30	47	47.3	-0.3
13	Stoke City	114	32	82	84.2	-2.2
14	West Ham United	122	43	79	81.2	-2.2
15	Swansea City	123	50	73	76.8	-3.8
16	Leicester City	119	42	77	82.6	-5.6
17	Bournemouth	113	46	67	72.8	-5.8
18	Southampton	76	31	45	51.3	-6.3
19	Manchester City	67	28	39	46.0	-7.0
20	Crystal Palace	118	46	72	80.4	-8.4

opportunities at a much higher rate. This is reflected in Table 6, where we can see how the increased rate of goals against is not explained by an increase in the number or quality of the shots

Table 9: Comparison of the distribution of unsuccessful shots conceded.

Team	Season	Shots	G	M	Blocked		Saved		Off target	
EPL Average	2015/16	12.3	1.2	11.1	4.3	39.1%	2.7	24.65%	4.0	36.3%
	2016/17	12.2	1.3	10.9	4.3	39.2%	2.7	25.05%	3.9	35.8%
Leicester City	2015/16	13.1	0.9	12.2	5.5	**45.4%**	2.5	20.36%	4.2	34.3%
	2016/17	13.7	1.6	12.1	4.5	**37.5%**	2.9	23.84%	4.7	38.7%

Table 10: Leicester City's 2016/17 goalkeepers side-by-side.

Player	N	Conceded	Saved	xS	Saved-xS
Kasper Schmeichel	85	27	58	57.9	0.1
Ron-Robert Zieler	34	15	19	24.7	-5.7

conceded, but by the effectiveness with which their opponents convert those chances. More specifically, the data shows defensive weaknesses against counter attacks, direct play and set plays (corner kicks and free kicks).

In terms of unsuccessful shots against, we saw in Section 2 how Leicester's field players were instrumental towards the team defensive performance by blocking rival opportunities. However, Table 9 shows this is no longer the case — the block rate has dropped almost 8 points and is now below the league average.

Finally, regarding goal-keeping performance, our expected save model also shows that the positive balance displayed in Table 3 has inverted, going from a situation where the goalkeeper prevented 4.6 expected goals during the course of the season to conceding 5.6 additional goals to what the model expected in just 25 games this campaign. One might think that this is due to Kasper Schmeichel underperforming this season, but this is not exactly the case: While last season he was the only one to take the goalkeeper role, this season he has shared that responsibility with teammate Zieler due to injury. As Table 10 indicates, his Saved-xS residual is now near zero so, although he has not been able to replicate the contributions he made last season, he is not underperforming either. Zieler, on the other hand, has not been able to match the expectation and has conceded 5.7 goals above that.

Obviously, another key factor is the lack of chances that Leicester are creating. Last year, the expected goals they created was

1.74 per game (66.1 for 38 games), compared to 1.13 (28.2 after 25 games) this season. As most of their chances are created via counter attacks, this is understandable as when they are behind the volume of chances they get via counter attack greatly diminishes as the opponents tend to sit back more, thus making them less susceptible to this type of chance. Having such strong priors in terms of both strategy and context enables accurate match prediction and strategy recommendation to occur (see next Section).

6 RECOMMENDATION ENGINE FOR OPPONENT SCOUTING

So far we have seen examples of the descriptive usefulness of the strategy plots introduced in the previous sections. The purpose of this final section is to show that these features also have capabilities in predicting future performance.

6.1 The Predictive Power of Strategic Features

The reason why analysts and coaches study the style of play of their opposition is, naturally, to try to anticipate what will happen in future encounters. With that in mind, we have carried out a prediction exercise to assess whether the strategic features can help in forecasting the way matches will play out.

To do so, we tackle the task of predicting the number of shots and goals that a team will score in a game. We compare the performance of four identical regression models (single-hidden-layer neural networks), trained 100 times on random splits (70% training, 30% validation) of the dataset, which consists of all games in the last 5 EPL seasons. Each game gives rise to two samples, one per team, so n=380x5x2=3800.

The models use the following input features: Model 1, our baseline reference, uses no inputs, and is therefore equivalent to using the average number of shots and goals in the training dataset as predictions. Model 2 uses a single binary input that indicates whether teams are playing home or away. Model 3 incorporates 14 additional inputs, each of which represent the average proportion of shots of each type that the corresponding team has made in the games that make the training dataset. Finally, model 4 adds 14 more inputs that again represent the average proportion of shots of each type taken, this time, by the opposition.

The results in Table 11 show that, unsurprisingly, the more information we give the model, the better it can predict the target indicators. The first performance bump comes when adding the venue indicator, however the largest one happens when we incorporate the strategy features for the team of interest, both in terms of number of goals and shots. Adding the style information of the opposition boosts the model once again.

This confirms that the strategy plots and the data powering them are not just a description of teams' preferences and habits, but that they are also correlated with their success or lack thereof on the pitch.

Table 11: Prediction performance (average mean squared error on the validation dataset over 100 training iterations) of the four regression models. The task was to predict the number of shots and goals that teams would score against their opponents.

	Inputs	Goals (MSE)	Shots (MSE)
Model 1	None	1.3528	26.5549
Model 2	Venue	1.3299	24.3382
Model 3	Venue + Strategy (team)	1.2485	21.5062
Model 4	Venue + Strategy (team & opposition)	1.2107	18.655

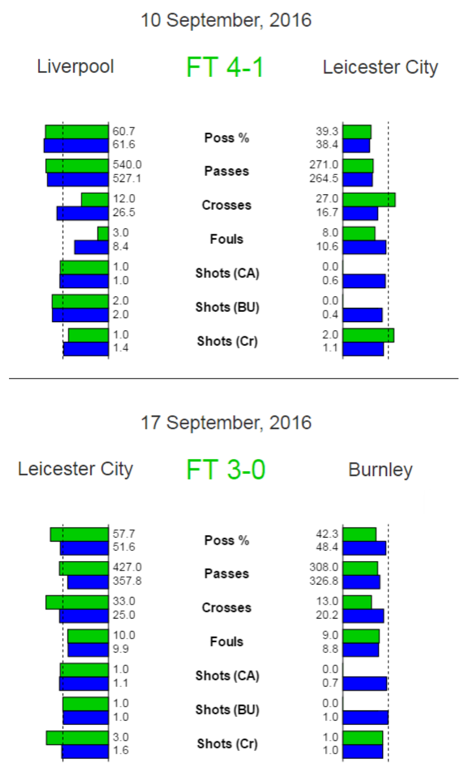

Figure 5: Examples of match outputs from the recommender system. Blue bars show the pre-match model predictions, green bars represent the actual values that took place. (a) Liverpool vs Leicester City (2016/17 season, week 4). (b) Leicester City vs Burnley (2016/17 season, week 5)

6.2 Predicting Future Strategy

We have just seen how teams' strategy can help get an idea of what will happen in terms of offensive volume (shots and goals). But, beyond that, what is probably more interesting from the point of view of pre-match preparation is how the game will play out from the point of view of strategy. Extending the approach in the previous section, we have developed a recommender system that estimates the expected shot production (by shot type) that teams will generate in an upcoming game.

Figure 5 shows examples of the types of outputs generated by our recommender system that highlight the importance of context. In the first game, Leicester City play away at against a tough opponent. Taking into account the style of both teams as well as the venue, the system predicts that it is likely a game that Liverpool will dominate from the point of view of possession, and where they will display a variety of shooting strategies. In contrast, the following week Leicester play at home against a weaker opponent, and in this case the estimates of the model are more favorable for Leicester in terms of game dominance and shot production.

7 SUMMARY

The 2015/16 EPL was an extraordinary campaign, and it will always stay in the memory of soccer fans. Looking at it from different perspectives, and using various machine learning techniques, this paper tries to uncover the keys to Leicester unimaginable success. The analysis of expected goal value in Section 2 makes apparent Leicester's defensive prowess, which was crucial to their victory. Their outfield players excelled in their defensive duties, and their goalkeeper Kasper Schmeichel topped the league in terms of above-expectation contribution.

Leicester were also very unique in their strategy. As Section 3 showed, their organized defensive structure allowed them to reduce the quality of their opponents' chances across the different scoring methods. Section 4 highlighted their disruptive game, embodied in the tireless efforts of N'Golo Kanté, that made them one of the most difficult teams to attack against. Offensively, they put in practice a style never before seen in a league champion, and they focused their shot production on the most dangerous strategies. However, in Section 5 we have shown that they have not performed to the same level in this current season.

The analysis tools we have used in this case study provide insightful information about strategy and efficiency in soccer. Furthermore, as Section 6 shows, their descriptive nature has a predictive byproduct that allows for their use as the core of a "recommendation engine" that can support professionals in the challenging task of pre-game preparation.

REFERENCES

[1] 2015. Why the English Premier League has been turned upside down. http://www.economist.com/blogs/gametheory/2015/12/competitive-balance-football. (19 December 2015). [Online; accessed 01-February-2017].

[2] Matias Grez. 2016. Champions Leicester City: 7 reasons why Foxes won Premier League title. http://edition.cnn.com/2016/05/03/football/leicester-city-champions-reasons-premier-league. (04 May 2016). [Online; accessed 01-February-2017].

[3] Michael Horton, Joachim Gudmundsson, Sanjay Chawla, and Joël Estephan. 2014. Classification of passes in football matches using spatiotemporal data. *arXiv preprint arXiv:1407.5093* (2014).

[4] Thom Lawrence. 2015. Expected saves. https://deepxg.com/2015/10/20/expected-saves. (20 October 2015). [Online; accessed 01-February-2017].

[5] P. Lucey, A. Bialkowski, M. Monfort, P. Carr, and I. Matthews. 2015. Quality vs Quantity: Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data. In *Proc. 9th Annual MIT Sloan Sports Analytics Conferenc.*

[6] Burr Settles. 2010. Active learning literature survey. *University of Wisconsin, Madison* 52, 55-66 (2010), 11.

[7] Peter Smith. 2016. Leicester City win Premier League: How they did it differently. http://www.skysports.com/football/news/30385/10267972/leicester-win-premier-league-how-they-did-it-differently. (03 May 2016). [Online; accessed 01-February-2017].

[8] Lukasz Szczepański and Ian McHale. 2016. Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179, 2 (2016), 513–533.

[9] Mark Taylor. 2014. How to calculate expected goal models. https://www.pinnacle.com/en/betting-articles/soccer/how-to-calculate-goal-models. (22 August 2014). [Online; accessed 01-February-2017].

[10] Colin Trainor. 2014. Goalkeepers: How repeatable are shot saving performances? http://statsbomb.com/2014/10/goalkeepers-how-repeatable-are-shot-saving-performances. (21 October 2014). [Online; accessed 01-February-2017].

[11] Yuji Yamamoto and Keiko Yokoyama. 2011. Common and unique network dynamics in football games. *PloS one* 6, 12 (2011), e29638.

A EXEMPLARS OF SHOT DICTIONARY

Visual depictions of all 15 types of shots in our shot dictionary are shown in Figures A.1 through A.7.

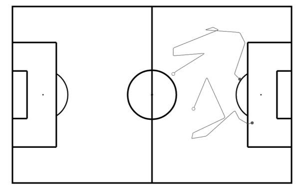

Figure A.1: Build Up: Build Up shots follow long ball possessions where the attacking team tries to attack their opponentfis goal through passing combinations.

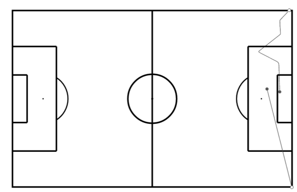

Figure A.2: Corner Kick: Shots that follow a corner kick. These can be long or short deliveries (bottom and top examples, respectively)

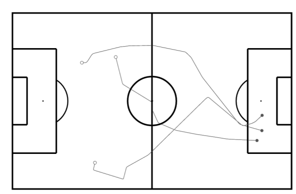

Figure A.3: Counter Attack: In the lead-up to the shot, the attacking team regains possession and quickly advances towards the rival goal.

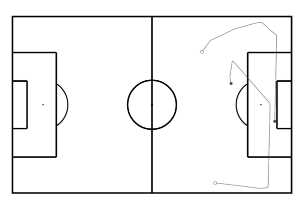

Figure A.4: Cross & From Cross: Shots in the Cross category immediately follow a crossed ball, whereas in shots labelled From Cross there are intermediate events between the cross and the shot.

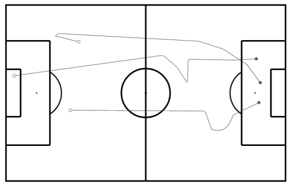

Figure A.5: Direct Play: Direct Play shots are preceded by a lonsg forward pass following the direction of attack.

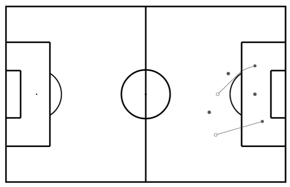

Figure A.6: Free Kick Shot, From Free Kick and Penalty Shot: Free Kick Shots are direct free kicks that are taken directly as a goal attempt, whereas shots From Free Kick happen after a free kick shot, or a pass or cross (direct or indirect). Penalty Shots are, as the name suggests, those following an infringement giving rise to a penalty kick.

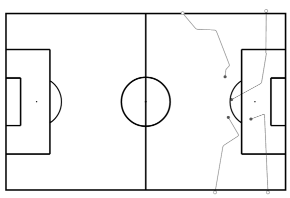

Figure A.7: In Throw In shots, the attempt follows a play resumption from the sideline.