

サッカーにおけるゴール期待値を使用したシミュレーションによる有用性の調査

経営学部経営学科

学籍番号：20161366

氏名 菅野瑠空

目次

目次

【第1章】はじめに	1
【1-1】研究の概要.....	1
【1-2】xGの概要.....	1
【1-3】研究の目的.....	2
【第2章】研究方法	2
【第3章】研究結果	4
【第4章】課題.....	7
【第5章】まとめ.....	8
【第6章】参考・引用文献.....	9

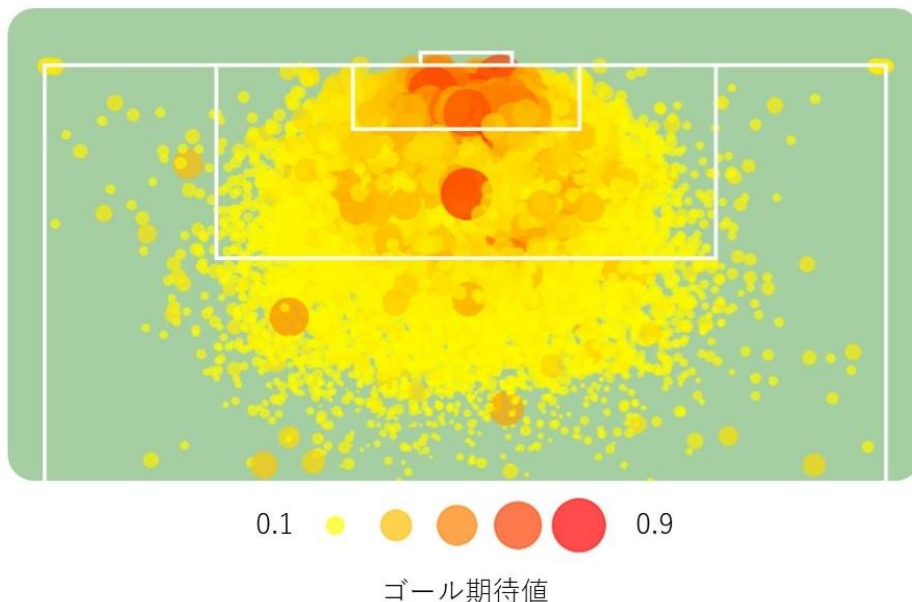
【第 1 章】はじめに

【1-1】研究の概要

本研究ではサッカーのデータ分析においてゴール期待値（以下 xG）と呼ばれるスタッツデータに焦点を当てて研究を行った。サッカーは非常に流動的なスポーツであり尚且つチームによって異なる複雑な戦術が取られるため、野球の総合指標である WAR のような他のスポーツでは存在している総合的に評価が出来るスタッツデータを生み出すことが難しいと考える。そういった中で近年扱われるようになってきた xG について注目を当てた。この xG というデータを用いてサッカーリーグのシミュレーションを行うことでそれぞれのチームの理想的な順位を示すことができ、その結果を実際の順位と照らし合わせることでその相違点や共通点から xG の有用性や課題を取り上げることが出来るのではないかと考えた。

【1-2】xG の概要

図 1 xG（ゴール期待値）のデータⁱ



まず xG とは日本のサッカーリーグである Jリーグとパートナーシップを結んでいるデータスタジアムが運営する Football Lab によると「ゴール期待値とは、『あるシュートチャンスが得点に結びつく確率』を 0~1 の範囲で表した指標であり、(中略) ゴール期待値はシュートの成功確率を表すので、値が高いほど得点が決まる可能性が高いシュートになります。」ⁱⁱと述べられおり、シュートを打った場面に対して選手のトラッキングデータやゴールに対しての角度、また、シュートに至るまでの直前のプレーのデータなどを過去のデータから一般的にそのシュートが入る確率を算出したものである。また、xG を可視化したものが図 1 である。この xG というデータは Football LAB では 2020 年から掲載されてい

るスタッツⁱⁱであり他にもスタッツデータの提供を行う企業である Opta 社は xG に関して 2017 年に解説動画を公開ⁱⁱⁱしている。また、xG という指標を生み出したのは footballista によると「この ExpG というコンセプトは、市井（しせい）の愛好家によって考案・発表され、インターネット上のコミュニティにおいて「オープンソース」的に開発が進められているところ。」^{iv}とあり、初出は公式的なデータではないことがわかる。これらのことから比較的近年扱われるようになったデータであり、正式なスタッツデータ分析企業から生み出されたものではないといえる。

【1-3】研究の目的

xG を用いて試合のシミュレーションを行う目的について、xG は上記の「xG の概要」で述べたようにこれまでのスタッツデータよりも比較的新しく生み出された項目であり、その有用性や重要性について議論が展開されているデータであると感じているため xG について焦点を当てて研究を行いたいと考えた。

また、この新たな観点から従来のシュート数やポゼッション率といったスタッツデータだけでは把握しきれなかった、シュートの質やチャンスクリエイティブといった試合のより深い理解を提供することができる重要性が高いデータであると考え。そのため xG を用いてシミュレーションを行い、実際の試合結果と比較することで xG がスタッツデータの中での重要度の高さを調査することが出来るのではないかと考えた。

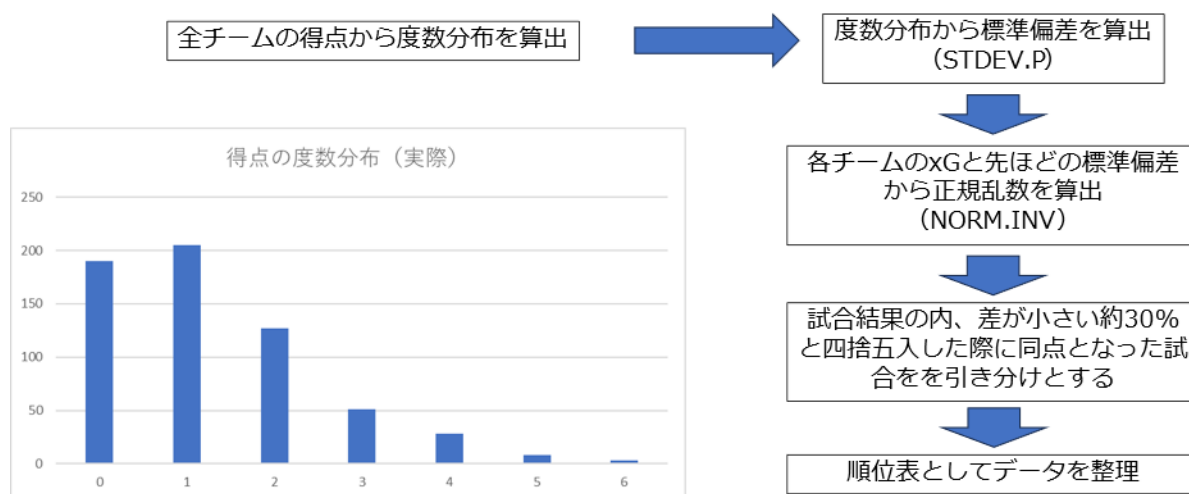
そのため今回の研究では xG のデータが高い有用性を持つと仮定する。具体的には xG と勝ち点との間に因果関係が存在し、シミュレーションを通じて得られる順位表は xG の高さに従い、降順に近い順位の並び順となるはずという仮説を立てた。これにより、サッカーの試合における xG の重要性がチームの実際の戦績や勝ち点にどのように影響するかについてより深く理解することが出来るようになる。

【第 2 章】研究方法

今回扱うデータは試合結果や基本的なスタッツを取り扱っている Football LAB から日本のサッカーリーグである Jリーグの 1 部リーグに当たる J1 リーグより、2022 年シーズンのデータ^v（チーム名、xG、勝利数、引き分け数、敗北数、得点数、失点数、得失点差）を用いた。なおチーム名はそれぞれ略称となっている。まず Excel にて上記データの全得点データから度数分布にまとめる。この得点の度数分布はポアソン分布の形になってしまうが正規分布として扱い標準偏差を算出する方法を取った。その後 18 チームごとにそれぞれの全試合平均の xG を平均とし、先ほどの標準偏差と組み合わせて正規乱数を発生させそのデータを全試合に当てはめた。この際、本来であれば正規分布でない度数分布を正規分布として扱っているため算出されたデータに本来であれば有り得ないマイナスの数値での結果が含ま

れてしまう場合がある。その場合は再度計算を行いプラスのデータになるまで計算を繰り返してそのデータを採用する形を取った。これを各 18 チーム 34 試合分対戦のシミュレーションを行った。数値に関して対戦結果の数値は整数の位まで四捨五入を行っている。試合の引き分けに関して Jリーグの公式レポートでは 2022 年シーズンの引き分け率は 31.7%とされていた^{vi}ため、シミュレーションされたデータの内対戦結果の数値で差が小さい約 30%の試合、または整数に四捨五入した際に同じ値になった対戦を引き分けとした。四捨五入したため引き分けにもかかわらず整数にした際に点差がつく対戦が生じてしまうが数値は素のまま引分けとして扱った。以上のデータを Jリーグの勝ち点のルールに則り、勝利の場合勝ち点を 3、引き分けの場合 1、敗北の場合 0 とし、加算して順位付けを行い勝ち点順の順位表としてまとめた。

図 2 シミュレーション方法



【第3章】研究結果

【3-1】研究結果の概要

表1 シミュレーション結果（1~17節）

1~17	横浜FM	川崎F	広島	鹿島	C大阪	FC東京	柏	名古屋	浦和	札幌	鳥栖	湘南	神戸	福岡	G大阪	京都	清水	磐田
横浜FM		1	4	0	3	1	2	1	1	1	1	2	3	1	4	3	2	1
川崎F	2		1	0	2	2	0	2	3	1	1	3	2	1	0	4	1	1
広島	2	1		1	0	0	1	2	1	1	0	0	3	0	2	1	2	1
鹿島	1	2	0		1	0	2	1	2	1	1	0	1	1	2	1	1	2
C大阪	4	1	3	2		2	0	1	3	0	2	2	0	0	0	4	1	2
FC東京	3	2	2	2	1		2	1	2	2	3	0	2	2	0	1	2	0
柏	0	1	0	0	2	0		2	0	0	3	1	2	1	4	1	0	0
名古屋	3	1	3	2	3	1	2		2	2	2	0	1	2	2	0	1	0
浦和	1	3	3	2	0	1	1	1		2	2	1	2	1	1	2	2	3
札幌	2	2	2	2	1	2	0	2	3		0	0	3	2	1	0	1	3
鳥栖	2	1	3	1	2	1	1	1	3	1		3	2	1	2	1	1	1
湘南	1	3	2	2	2	1	2	1	2	3	1		1	1	0	1	2	3
神戸	1	1	2	4	1	0	1	2	2	3	1	0		1	0	0	3	1
福岡	3	2	1	1	0	3	0	1	3	3	1	1	2		0	2	2	1
G大阪	3	2	2	2	2	0	1	2	2	3	1	1	1	1		4	2	1
京都	4	0	0	3	0	2	2	1	2	5	1	1	0	1	2		2	2
清水	2	0	1	1	1	1	1	2	0	1	2	1	1	2	2	2		2
磐田	2	2	4	3	0	1	2	1	4	2	3	1	3	1	2	1	2	

表2 シミュレーション結果（18~34節）

18~34	横浜FM	川崎F	広島	鹿島	C大阪	FC東京	柏	名古屋	浦和	札幌	鳥栖	湘南	神戸	福岡	G大阪	京都	清水	磐田
横浜FM		1	0	2	0	1	2	2	1	0	3	2	3	2	3	1	2	3
川崎F	4		0	1	3	0	3	4	3	4	1	2	4	2	4	2	0	2
広島	2	4		1	1	2	2	2	1	1	0	1	1	1	1	0	2	1
鹿島	1	2	1		2	1	1	1	3	2	2	1	0	2	0	1	3	1
C大阪	2	0	1	2		2	1	1	1	1	2	2	3	1	3	3	0	4
FC東京	3	4	1	1	1		2	1	1	2	1	3	1	2	2	2	4	1
柏	1	1	1	0	1	2		3	0	3	1	1	2	1	2	3	2	4
名古屋	3	1	1	1	2	0	0		2	2	5	2	2	1	1	1	0	2
浦和	4	2	1	0	1	0	1	1		1	1	1	1	2	3	2	1	2
札幌	2	2	3	1	2	3	1	3	1		4	2	4	3	1	2	1	1
鳥栖	2	3	0	2	4	1	2	2	1	1		0	2	1	3	3	1	0
湘南	1	1	2	2	1	1	1	2	3	2	3		1	1	0	3	1	1
神戸	1	0	1	3	1	2	1	2	2	2	2	1		1	4	2	1	3
福岡	1	1	0	1	0	1	1	1	2	1	3	0	2		0	2	1	1
G大阪	2	2	2	1	2	1	3	3	2	0	3	2	2	2		1	0	1
京都	4	3	2	1	0	1	2	2	1	1	2	2	1	2	1		0	2
清水	3	1	0	0	0	2	2	1	1	2	2	2	3	2	1	1		1
磐田	2	2	2	2	1	2	1	2	1	0	0	1	1	2	1	2	4	

*列が得点、行が失点の数値

*配列は実際の順位に沿った順序

*黄色の塗りつぶしは引き分けの
対戦結果であることを示す

表 3 順位表 (シミュレーション)

順位 (実際)	チーム名	勝点	xG	xG順位	勝	分	敗	得点	失点	得失
1	横浜FM	57	1.849	1	17	6	11	74	59	15
9	浦和	55	1.567	2	14	13	7	61	52	9
7	柏	54	1.156	12	15	9	10	46	45	1
13	神戸	54	1.369	6	15	9	10	62	52	10
3	広島	50	1.37	5	13	11	10	51	41	10
14	福岡	49	1.056	17	12	13	9	47	44	3
8	名古屋	48	1.232	8	12	12	10	57	53	4
16	京都	47	1.159	10	12	11	11	59	55	4
4	鹿島	46	1.152	13	10	16	8	49	43	6
17	清水	45	1.183	9	10	15	9	50	46	4
2	川崎F	43	1.411	3	12	7	15	55	65	-10
11	鳥栖	43	1.084	15	9	16	9	60	55	5
5	C大阪	39	1.253	7	9	12	13	43	56	-13
6	FC東京	38	1.157	11	10	8	16	40	59	-19
15	G大阪	38	1.03	18	8	14	12	54	59	-5
18	磐田	37	1.11	14	9	10	15	54	60	-6
10	札幌	36	1.375	4	9	9	16	56	62	-6
12	湘南	36	1.064	16	8	12	14	42	54	-12

表 4 順位表 (実際)

順位 (実際)	チーム名	勝点	xG	xG順位	勝	分	敗	得点	失点	得失
1	横浜FM	68	1.849	1	20	8	6	70	35	35
2	川崎F	66	1.411	3	20	6	8	65	42	23
3	広島	55	1.37	5	15	10	9	52	41	11
4	鹿島	52	1.152	13	13	13	8	47	42	5
5	C大阪	51	1.253	7	13	12	9	46	40	6
6	FC東京	49	1.157	11	14	7	13	46	43	3
7	柏	47	1.156	12	13	8	13	43	44	-1
8	名古屋	46	1.232	8	11	13	10	30	35	-5
9	浦和	45	1.567	2	10	15	9	48	39	9
10	札幌	45	1.375	4	11	12	11	45	55	-10
11	鳥栖	42	1.084	15	9	15	10	45	44	1
12	湘南	41	1.064	16	10	11	13	31	39	-8
13	神戸	40	1.369	6	11	7	16	35	41	-6
14	福岡	38	1.056	17	9	11	14	29	38	-9
15	G大阪	37	1.03	18	9	10	15	33	44	-11
16	京都	36	1.159	10	8	12	14	30	38	-8
17	清水	33	1.183	9	7	12	15	44	54	-10
18	磐田	30	1.11	14	6	12	16	32	57	-25

表 5 得点順 (シミュレーション)

順位 (シミュ)	チーム名	勝点	xG	xG順位	勝	分	敗	得点	失点	得失
1	横浜FM	57	1.849	1	17	6	11	74	59	15
4	神戸	54	1.369	6	15	9	10	62	52	10
2	浦和	55	1.567	2	14	13	7	61	52	9
12	鳥栖	43	1.084	15	9	16	9	60	55	5
8	京都	47	1.159	10	12	11	11	59	55	4
7	名古屋	48	1.232	8	12	12	10	57	53	4
17	札幌	36	1.375	4	9	9	16	56	62	-6
11	川崎F	43	1.411	3	12	7	15	55	65	-10
15	G大阪	38	1.03	18	8	14	12	54	59	-5
16	磐田	37	1.11	14	9	10	15	54	60	-6
5	広島	50	1.37	5	13	11	10	51	41	10
10	清水	45	1.183	9	10	15	9	50	46	4
9	鹿島	46	1.152	13	10	16	8	49	43	6
6	福岡	49	1.056	17	12	13	9	47	44	3
3	柏	54	1.156	12	15	9	10	46	45	1
13	C大阪	39	1.253	7	9	12	13	43	56	-13
18	湘南	36	1.064	16	8	12	14	42	54	-12
14	FC東京	38	1.157	11	10	8	16	40	59	-19

【3-2】 研究結果の考察

上記が今回のシミュレーション結果である。想定では勝ち点による順位はxGに従って降順のような形になると仮定しており、実際にxGが最も高い横浜FMがシミュレーションでも一位となっている。それだけでなくxGにおいて2位と非常に高い順位にもかかわらず本来の順位では9位と結果が芳しくなかった浦和もシミュレーションでは2位と上位に名を連ねている。しかし、続いて3位となったのは柏でありxGでは12位で下位に沈むチームとなっている。そこで全体として比較したところ、シミュレーション結果上位半分の9チームの内xGの順位でも上位半分に入っているチームは5チームと半数以上がどちらも上位に来ているが仮定していたほどxGの高さに倣った順位になっているとは言えない。

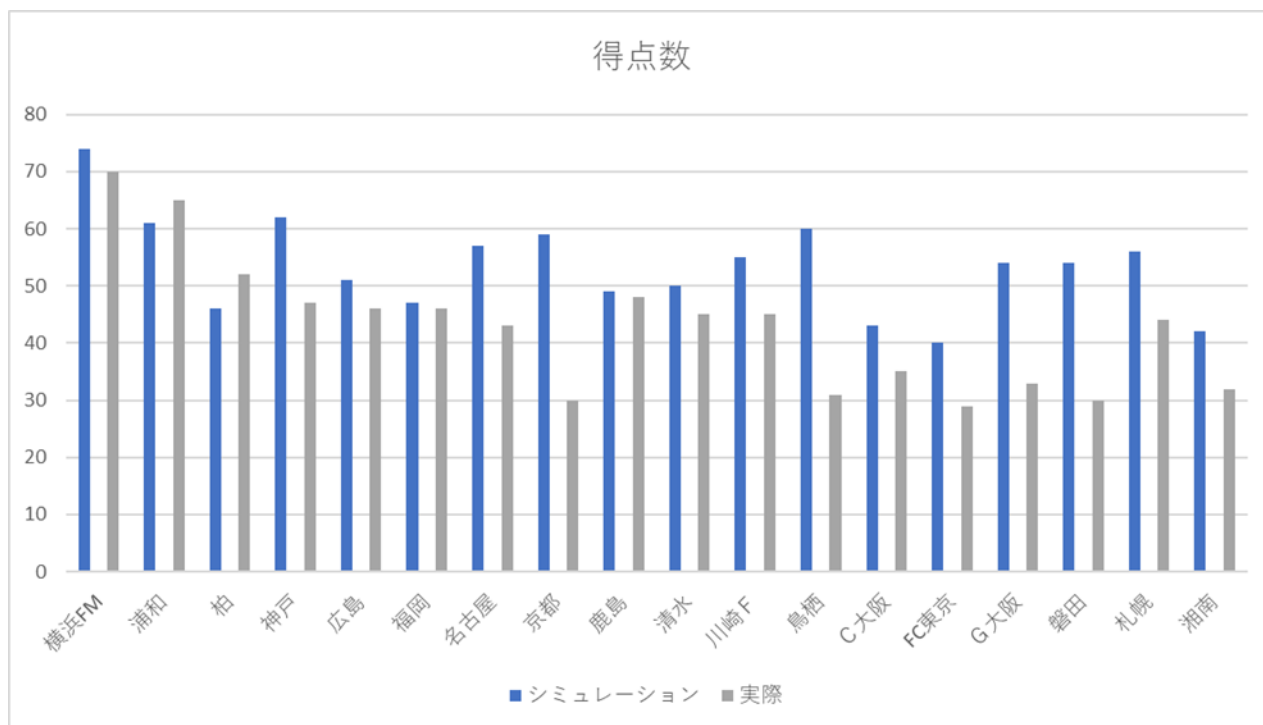
他に、上記でxGが低いにもかかわらず柏が3位という結果になっていると述べたが注目すべき点として上位の他チームと比較して得点数が46点と劣っている部分である。そこでより深く考察するために表3において得点数順に順位をソートし表示している。表3の上位9チームの内、xGの順位でも上位9位に来ているチームは6チームと勝ち点での順位より1チーム多く、3分の2が得点数とxGの両方において上位であり、わずかではあるが勝ち点による順位よりも得点数の順位の方がよりxGを反映していると考えられる。このことからxGを重視した場合、より顕著に結果が直結するデータは勝ち点や勝

利ではなく得点であり、xG を重要視することが必ずしも勝利という結果に直結するわけではないと考える。

また、得失点差に関して実際の試合結果では横浜 FM や川崎 F といったチームの得失点差がプラスに大きくなっており、突出しているがシミュレーションによる試合結果ではそういった突き出た得失点差が存在していないように見える。これは失点に関するデータがシミュレーションに含まれていないため、本来であればチームによって差が出るはずの失点数が均されていることから起きた結果であると考えられる。

他にも得点数を実際の数値と比較すると、シミュレーションによる結果側が高い傾向となったことが表 6 から読み取れる。このような結果となった理由として考えられるのは、シミュレーション時に全体の得点分布をもとにして計算を行ったが、本来であればチームごとに得点分布は異なるという点を考慮しなかった事、2つ目の理由として、上記でも述べたように失点に関するデータを使用しなかったためチームの特徴として守りの堅い、といったチームがシミュレーション上において存在しなくなり、得点数が増加したということが要因となっていると考える。

図 6 得点数の比較



【第 4 章】課題

本研究の課題として、xG だけでなく失点期待値をシミュレーションに入れ込む必要性があったと考える。その理由として xG のみでのシミュレーションの場合、評価されるのはチームの攻撃能力の部分のみで守備能力は一切評価対象に含まれないためである。そのため実際の順位表で失点数の少なかった

名古屋といったチームは評価をされずにシミュレーションをされてしまった。以上のことから再度研究を行う際には失点期待値を考慮することによってより複雑かつ効果的な研究を行うことが出来るのではないかと考える。

また、今回引き分けの定義として試合結果の内差が小さい30%の試合または整数に四捨五入した際に対戦したチームそれぞれが同じ値になった試合を引き分けと置いた。そのため想定よりも多く引き分けが出てしまった。したがって次回研究を行う際はまず引き分けの数値は複数年の割合から定義を行い、なおかつ四捨五入した際に同じ値になってしまい意図せず引き分けとした試合をなくすために数値の繰り下げや繰り上げなど整数化に工夫が必要だと感じた。

他にも得点の度数分布がポアソン分布だったにもかかわらず正規分布として正規乱数を発生させたことからマイナスの値がデータに含まれるようになってしまったためより正確なデータを求めるのであればポアソン分布での乱数の生成方法によって計算を行うべきだったと感じた。

また、今回時間の都合上行うことが困難であったが単一のリーグで複数年のスタッツデータを用いてシミュレーションを行うことや、データが存在すれば様々なリーグを取り上げてシミュレーションし、その結果の共通点からxGについて新たな知見を探ることや相違点からそれぞれのリーグの特徴を見つけ出すことが出来たのではないかと考えた。

【第5章】まとめ

本研究ではサッカーのスタッツデータの一つであるxGを使用してシミュレーションを行い、分析を行った。今回のシミュレーションではあくまで攻撃的な部分をピックアップし利用したため守備的な部分でデータが十分ではなかったことが前提ではあるが、今回の結果からはxGを重要視することは勝利に直接的に関係するとは言い切れないが、得点といった部分で間接的に関係するという推測に至った。

今後の展望としてxGと実際の得点結果がそれぞれチームごとにどのように違っているかの調査や、実際の試合結果とシミュレーション結果に差異が生じた場合にその理由を推測することでxGを活用する上で指針になる可能性があると考え。他にも失点期待値をデータに含めることや、失点期待値を含めたシミュレーションを行うことでよりスタッツデータを深く掘り下げたいと考える。また、研究を続けることでxGが過去のトラッキングデータやシュートの種類といったデータを利用して生みだされたように、xGを利用して新たな指標となるスタッツデータを作り出す可能性を含んでいるとも考える。

【第 6 章】 参考・引用文献

ⁱ “フットボールラボとは”.Football LAB.

<https://www.football-lab.jp/pages/about/>. (閲覧日 2023-12-26)

ⁱⁱ “ゴール期待値とは”.Football LAB.

https://www.football-lab.jp/pages/expected_goal/, (閲覧日 2023-12-22)

ⁱⁱⁱ “Opta Expected Goals”. Opta Analyst_YouTube.

https://www.youtube.com/watch?v=w7zPZsLGK18&ab_channel=OptaAnalyst, (閲覧日 2023-12-22)

^{iv} “進化型サッカー評論④ゴール期待値「ExpG」”.footballista.

<https://www.footballista.jp/special/31762> , (閲覧日 2023-12-26)

^v “2022 J1 順位表 Summary”. Football LAB.

https://www.football-lab.jp/summary/team_ranking/j1/?year=2022 (閲覧日 2023-12-26)

^{vi} “J.LEAGUE J STATS REPORT 2022 (概要) “. P4.【公式】Jリーグ公式サイト (J.LEAGUE.jp)

https://www.jleague.jp/img/pdf/REPORT_2022_1 (閲覧日 2023-12-26)

指導教員からの講評：

本研究は、公開されている各チームの xG に基づき、2022 シーズンの対戦シミュレーションを行ったもので、今回の結果としてはラフなものですが、今後にもつながる大変興味深い研究成果でした。論文の質も高く、丁寧にまとめられていると思います。

今後の方向性としては、課題として挙げられているように、データの拡充と、失点を含めたシミュレーションが可能性としてはありそうです。