

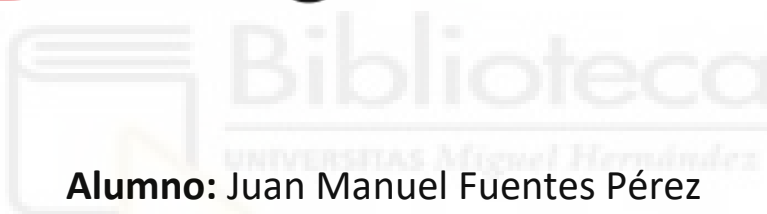


Trabajo Fin de Grado

**Revisión bibliográfica sobre la aplicación del
“modelo de Goles Esperados (xG)” para el análisis
de rendimiento en fútbol**



UNIVERSITAS
Miguel Hernández



Alumno: Juan Manuel Fuentes Pérez

Tutor: Tomás Urban Infantes

Grado en Ciencias de la Actividad Física y del Deporte
Universidad Miguel Hernández de Elche

2021-2022

ÍNDICE PAGINADO:

RESUMEN Y PALABRAS CLAVE	3
1. CONTEXTUALIZACIÓN	3
2. PROCEDIMIENTO DE REVISIÓN (METODOLOGÍA)	4
2.1. MATERIALES Y MÉTODOS.....	4
2.1.1 FUENTES DE INFORMACIÓN.....	4
2.1.2 ESTRATEGIA DE BÚSQUEDA.....	4
2.1.3 SELECCIÓN DE ESTUDIOS.....	5
2.1.4 EXTRACCIÓN DE DATOS.....	5
2.2 RESULTADOS.....	5
3. REVISIÓN BIBLIOGRÁFICA (DESARROLLO)	6
3.1 DESARROLLO Y APLICACIÓN DEL “MODELO DE GOLES ESPERADOS (xG)”	7
3.2 ANÁLISIS DE RENDIMIENTO ESTRATÉGICO-TÁCTICO.....	8
3.3 MITIGACIÓN DE SEGOS DE RESULTADO EN EL ANÁLISIS DE RENDIMIENTO.....	9
4. DISCUSIÓN	11
4.1 LIMITACIONES.....	13
4.2 CONCLUSIONES.....	13
5. PROPUESTA DE INTERVENCIÓN	14
6. BIBLIOGRAFÍA	15
7. ANEXOS	17

RESUMEN

En la actualidad, el Big Data y los últimos avances en computación han revolucionado los métodos de análisis de rendimiento en aquellos deportes de naturaleza más caótica. El objetivo de este trabajo es examinar sistemáticamente las aplicaciones del “modelo de Goles Esperados (xG)” para el análisis de rendimiento en fútbol. Se realizó una búsqueda en las bases de datos PubMed, SPORTDiscus y Scopus, y se llevó a cabo una revisión sistemática de la literatura según las directrices PRISMA. Los resultados obtenidos muestran diferentes enfoques en el desarrollo del modelo xG; así como sus aplicaciones para apoyar objetivamente la toma de decisiones analizando la producción ofensiva y defensiva de equipos y jugadores, el planteamiento estratégico-táctico, y mitigando la influencia de los factores aleatorios sobre la evaluación de los resultados. Las métricas xG pueden ser de gran utilidad para las labores de scouting, el estudio del juego en profundidad, y la predicción de resultados futuros; abriendo nuevas oportunidades de investigación en el área del análisis de rendimiento en fútbol.

Palabras clave: Fútbol, big data, análisis de rendimiento, scouting, goles esperados.

1. CONTEXTUALIZACIÓN

El reconocimiento del uso de análisis estadístico en el deporte se remonta hasta 1859 cuando Henry Chadwick desarrolló el “box score” o cuadro de estadísticas en béisbol (Eaves, 2015). El análisis notacional empezó a usarse como método objetivo para registrar y cuantificar los eventos clave que determinan el rendimiento deportivo, de un modo consistente y fiable (Morgulev et al., 2018). Este análisis permite a los entrenadores y agentes involucrados en el proceso de toma de decisiones de las organizaciones deportivas evaluar objetivamente el rendimiento competitivo y, por lo tanto, orientar su mejora (Morgulev et al., 2018).

El cricket o el béisbol son deportes más susceptibles de ser analizados mediante el uso del análisis notacional dada su naturaleza y reglamento, ya que pueden ser descompuestos fácilmente en eventos discretos claramente diferenciados; al contrario que ocurre en deportes como el fútbol o el hockey en los que su naturaleza continúa y dinámica, y con interacciones más complejas entre jugadores, dificultan enormemente la labor de registro y análisis significativo de los datos (Szymanski, 2020).

Históricamente, el ejemplo más claro de la aplicación del análisis notacional al fútbol ha sido orientado al estudio sobre dónde es más probable que los jugadores rivales lancen un penalti, información que los analistas ofrecían al portero previamente al partido, basándose en el registro histórico de lanzamientos de penalti de los oponentes (Memmert et al., 2013).

En los últimos 20 años, los avances tecnológicos en computación, monitorización y en la retransmisión de los eventos deportivos han permitido la recolección y el modelado estadístico de enormes conjuntos de datos sobre la localización y el movimiento de los jugadores en el campo, parámetros físicos a lo largo de la temporada, pases, ángulos de disparo a portería, etc.; incluso consiguiendo descomponer implacablemente aquellos deportes más complejos de analizar como el fútbol en eventos más sencillos, para extraer información que sirva de utilidad al cuerpo técnico, equipo directivo, periodistas, aficionados, etc.

Este tipo de avances, han dado lugar al concepto actual de “Análítica Deportiva” o “Big Data Deportivo”, que permite descubrir patrones entre grandes conjuntos de datos de un modo en el que la capacidad computacional no permitía hacerlo hasta ahora, y usar modelos estadísticos predictivos para apoyarnos en el proceso de toma de decisiones basándonos en los datos (Morgulev et al., 2018; Szymanski, 2020).

En fútbol, Lago-Peñas et al., 2010, 2011 determinaron que tanto el número de disparos como el número de disparos a puerta, son dos de las principales métricas estadísticas que mejor discriminan entre los equipos ganadores y perdedores en competiciones profesionales (España LaLiga y UEFA Champions League).

Dada la naturaleza poco anotadora de este deporte, tiene sentido que se utilicen estos datos como indicadores clave de rendimiento, teniendo en cuenta el impacto que un gol supone en un partido de fútbol comparado con otros deportes de naturaleza más anotadora (una canasta en baloncesto, una carrera en béisbol, etc.). Sin embargo, no todos los disparos son creados de la misma manera, y su calidad difiere significativamente en función de la situación. Además, sólo el 1% de todas las jugadas de ataque y 10% de los tiros acaban en gol (Pollard & Reep, 1997; Tenga et al., 2010), por lo que evaluar el rendimiento basándonos sólo en esta métrica binaria (gol o no gol) provoca que se antepongan los resultados al proceso y se pase por alto información relevante para el análisis (Anzer & Bauer, 2021).

En el fútbol profesional, constantemente se toman decisiones con consecuencias millonarias basadas en la lógica de que los resultados de los partidos reflejan con fidelidad el rendimiento real en el campo (Brechot & Flepp, 2020), infraestimando la influencia del azar sobre estos y exhibiendo de este modo un evidente sesgo de resultado (Gauriot & Page, 2019). Consecuentemente, los managers suelen fallar a la hora de hacer ajustes necesarios tras victorias fortuitas y actúan excesivamente tras derrotas desafortunadas (Lefgren et al., 2015).

La idea principal detrás de los “Goles Esperados” (de aquí en adelante, “xG”), es dar más importancia al proceso de producción real en el terreno de juego que a los resultados en la evaluación del rendimiento. Los tiros ocurren mucho más frecuentemente que los goles y por lo tanto son menos propensos a la influencia de los factores aleatorios. Además, conforman el último paso del proceso de producción antes de marcar; y un planteamiento en el que se busque maximizar las buenas ocasiones generadas y reducirlas para el equipo rival, es intuitivo futbolísticamente (Brechot & Flepp, 2020).

El “modelo de Goles Esperados (xG)” permite cuantificar la calidad de cada tiro, atribuyéndole a cada uno su probabilidad de materializarse como gol (valores entre 0 y 1), en base a los diferentes factores que describen el disparo (Anzer & Bauer, 2021; Rathke, 2017)

El objetivo de esta revisión bibliográfica es examinar la aplicación del “modelo xG” para el análisis de rendimiento en fútbol.

2. PROCEDIMIENTO DE REVISIÓN (METODOLOGÍA)

2.1 MATERIALES Y MÉTODOS

Se realizó una revisión sistemática de la literatura disponible según las directrices PRISMA (Preferred Reporting Items for Systematic reviews and Meta-analyses) (Moher et al., 2015).

2.1.1 FUENTES DE INFORMACIÓN

Se llevó a cabo una búsqueda bibliográfica hasta el 26 de marzo de 2022 para identificar estudios que informaran del uso del modelo xG para el análisis de rendimiento en fútbol. Exactamente, fueron consultadas las siguientes bases de datos electrónicas: PubMed, SPORTDiscus y Scopus.

2.1.2 ESTRATEGIA DE BÚSQUEDA

Se buscaron los títulos y/o resúmenes de todos los artículos usando la combinación de los siguientes operadores booleanos en PubMed y SPORTDiscus: (“expected goals” OR xG) AND (football OR soccer) NOT (“eye-xg” OR “tracking-xg”); y en Scopus: TITLE-ABS-KEY (“expected goals” OR xG) AND TITLE-ABS-KEY (football OR soccer) AND NOT TITLE-ABS-KEY (“eye-xg” OR “tracking-xg”).

2.1.3 SELECCIÓN DE ESTUDIOS

Para ser tenidos en cuenta en esta revisión, los estudios debían **analizar el rendimiento deportivo en fútbol** aplicando el “**modelo de goles esperados (xG)**”, y cumplir con los siguientes criterios de inclusión y exclusión:

Criterios de inclusión:

- Publicado en los últimos 10 años.
- Publicación de texto completo disponible en inglés.
- Publicado en una revista científica.
- Datos recogidos en fútbol.
- Datos recogidos en ligas profesionales.
- Datos recogidos en partido.
- Datos disponibles de más de un equipo.
- Datos disponibles de dos o más jugadores.

Criterios de exclusión:

- No disponible texto completo en inglés.
- Datos recogidos en fútbol sala / americano / australiano.
- Sólo datos disponibles de un único equipo.
- Presenta datos de menos de dos jugadores.
- Datos recogidos en entrenamiento.

Se realizó una primera selección de artículos en base al título y resumen. Cualquier estudio que claramente no cumpliera con los criterios de inclusión, fue excluido en esta fase. Cuando no era posible tomar una decisión segura en base al título y resumen, la publicación se incluía para el análisis posterior del texto completo. Finalmente, la elegibilidad para la inclusión definitiva de los estudios restantes en la revisión se evaluó tras realizar el análisis del texto completo.

2.1.4 EXTRACCIÓN DE DATOS

La siguiente información fue extraída de los artículos incluidos: (1) autores, (2) objetivos, (3) muestra, (4) métodos, (5) resultados y (6) conclusiones.

2.2 RESULTADOS

La búsqueda inicial en las tres bases de datos arrojó un total de 28 registros para su inclusión (véase “Identificación” en la Figura 1). Tras eliminar los duplicados, y filtrar los registros en base a título y resumen, quedaron excluidos 16 estudios. Como resultado, se seleccionaron un total de 12 registros que fueron considerados para la evaluación del texto completo (véase “Cribado” en la Figura 1). Tras la evaluación del texto completo, se excluyeron 6 registros porque no cumplían los criterios de inclusión (véase “Idoneidad” en la Figura 1). Los 6 estudios restantes fueron incluidos para el análisis de la revisión. A continuación, se describirán los resultados de la revisión sistemática de la literatura.

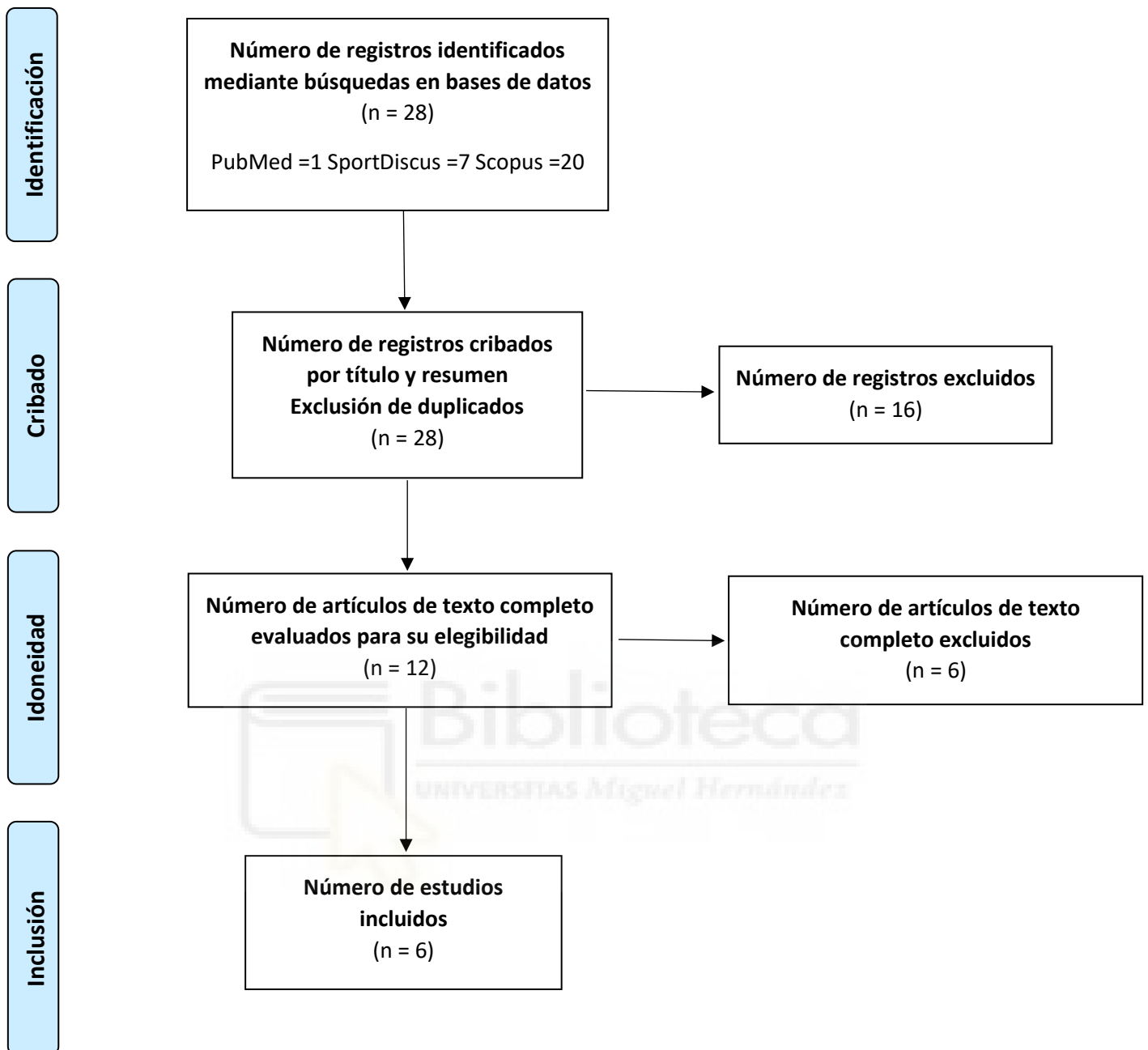


Figura 1. Diagrama de flujo de la búsqueda sistemática de la literatura (conforme a las directrices PRISMA)

3. REVISIÓN BIBLIOGRÁFICA (DESARROLLO)

Para una mejor estructuración y análisis de la información, los artículos se clasificaron según su objeto de estudio en las siguientes tres categorías:

- Desarrollo y aplicación del “Modelo de Goles Esperados (xG)”.
- Análisis de rendimiento estratégico-táctico.
- Mitigación de sesgos de resultado en el análisis de rendimiento.

3.1 DESARROLLO Y APLICACIÓN DEL “MODELO DE GOLES ESPERADOS (xG)”

Autor/es	Objetivo	Muestra	Métodos	Resultados	Conclusiones
(Rathke, 2017)	<ul style="list-style-type: none"> - Calcular modelo xG utilizando distancia entre el punto de disparo y portería. - Usar xG para calcular eficiencia de tiro. 	18.218 chuts Premier League y Bundesliga 2012/2013 – NO córner, centro, libre directo, gol propia puerta, penalti, 1ª mitad campo.	<ul style="list-style-type: none"> - Estudia ligas por separado. - Distribuye tiros en 8 zonas del campo (X = distancia portería, Y = ángulo portería) (Anexo 1). - Probabilidad xG zona = % ('nº goles zona / nº tiros zona'). - xG totales zona = 'nº tiros zona' x 'probabilidad xG zona'. - xG equipo → Sumar 'xG totales todas las zonas' del equipo. - Mismo procedimiento para calcular xG en contra. - Eficiencia tiro = nº goles / xG. 	<ul style="list-style-type: none"> - El modelo xG que tiene en cuenta distancia y ángulo a portería, es más preciso que el que sólo tiene en cuenta la distancia. - Bondad de ajuste muy elevada entre goles reales marcados y xG. En equipos (Bundesliga R2= 0.81 Premier League R2= 0.84) y jugadores (R2= 0.81 ambas ligas). - En ambas competiciones, equipos mejor clasificados rindieron por encima de sus goles esperados y los peor clasificados rindieron por debajo. 	<ul style="list-style-type: none"> - Estima el rendimiento goleador con relativa precisión. - Menos preciso cuando estiman xG de equipos mejor/peor clasificados debido a la habilidad de sus jugadores. - Podemos identificar jugadores que rinden consistentemente mejor, peor o similar a sus goles esperados. - Los estudios de la eficiencia de tiro dependen del modelo xG. - El estudio de xG puede aplicarse para mejorar la comprensión del juego y el análisis táctico.
(Anzer & Bauer, 2021)	<ul style="list-style-type: none"> - Desarrollar modelo xG que utilice datos posicionales y de evento, para cuantificar la calidad de cualquier disparo. - Introducir la aplicación del modelo al análisis de rendimiento. 	<p><u>Desarrollo modelo:</u> 105.627 tiros Bundesliga y Bundesliga 2 2013/2014 – 2019/2020.</p> <p>Sistema visión artificial 25 Hz: posición 22 jugadores, árbitro y balón respecto a límites del campo.</p>	<p><u>Desarrollo del modelo:</u></p> <ul style="list-style-type: none"> - Diseñan algoritmo de sincronización de datos de posicionamiento y evento. - Desarrollan modelo xG comparando 5 modelos de aprendizaje automático supervisado; computados con factores que influyen en la probabilidad de marcar gol (definidos por analistas tácticos profesionales – Anexo 2). - Aplican predicción con modelo xG definitivo a todos los disparos 3 primeras jornadas 2020/2021 Bundesliga 1 y 2 (1.357 disparos). <p><u>Análisis rendimiento:</u></p> <ul style="list-style-type: none"> - Comparan correlación entre 'ratio goles futuro' (goles marcados – goles concedidos) y varias métricas agregadas (ratio xG, ratio disparos a puerta, ratio total de disparos, ratio 	<p><u>Desarrollo del modelo:</u></p> <ul style="list-style-type: none"> - Desfase absoluto promedio del algoritmo de sincronización respecto a 'Gold Standard' 0,23 (±0,49) s). - Modelo xG final seleccionado (predice mejor) → modelo de 'potenciación de gradiente' (regresor y clasificador estadístico) que combinaba 3 submodelos diferenciados para chuts con la pierna, remates de cabeza y libres directos → error de estimación promedio respecto a muestra completa 0.0928. - De los 1.357 tiros, 150 acabaron en gol. El modelo predijo un valor xG agregado de 151,6. 	<p><u>Desarrollo del modelo:</u></p> <ul style="list-style-type: none"> - Algoritmo de sincronización supone mejora sustancial para análisis de disparos → Permite sincronizar grandes muestras de datos posicionales y de evento de manera rápida y relativamente precisa. - Modelo xG final predice resultado de disparos mejor que cualquier otro modelo xG revisado hasta la fecha. <p><u>Análisis rendimiento:</u></p> <ul style="list-style-type: none"> - Incluir datos de posicionamiento en modelo xG permite análisis más granular de la influencia de los factores concernientes al disparo en la probabilidad de marcar. - Permite reducir aleatoriedad de los resultados en análisis de rendimiento, diferenciando entre victorias merecidas (mayor valor agregado xG que rival → mayor nº ocasiones o de mayor calidad), y no merecidas.

		<p><u>Análisis rendimiento:</u> 5.450 tiros Bundesliga 2019/2020.</p>	<p>goles, puntos por partido), tras un cierto día de partido.</p> <ul style="list-style-type: none"> - Analizan la influencia de los factores concernientes al disparo en la probabilidad de gol predicha. - Comparación diferencia goles marcados y concedidos, con diferencia valores xG agregados computados por el modelo. Para cada equipo durante la temporada y para cada partido por separado. - Valores xG agregados de cada equipo durante el transcurso de un partido. - Estudio comportamiento de disparo equipos y jugadores (nº y calidad promedio de disparos por partido) 	<p><u>Análisis rendimiento:</u></p> <ul style="list-style-type: none"> - Consistentemente ratio xG predecía ratio goles futuros mejor que cualquier otra métrica (especialmente a partir de la jornada 10). - 73,3% todos los partidos (no empates), ganador tenía mayor valor xG. Sólo en 56,2% partidos, ganador tenía más disparos. - Factores con mayor impacto en probabilidad de gol del disparo son: ubicación disparo (distancia y ángulo), distancia portero a portería. 	<ul style="list-style-type: none"> - Permite estudiar evolución comportamiento de disparo de un equipo durante un partido → Adaptar estrategia defensiva. - Estudiar contribución ofensiva jugador de modo más granular que con métricas tradicionales. - Estudiar eficiencia de disparo identificando jugadores que rinden consistentemente mejor, peor o similar a sus xG. -Asumiendo disparos “simulados” en distintos puntos de una misma jugada y su valor xG, podemos cuantificar la contribución de cada jugador a la probabilidad final de marcar, y evaluar toma de decisiones. - Valores xG sirven de base para calcular otras métricas (ej.: asistencias esperadas).
--	--	---	---	--	--

3.2 ANÁLISIS DE RENDIMIENTO ESTRATÉGICO-TÁCTICO

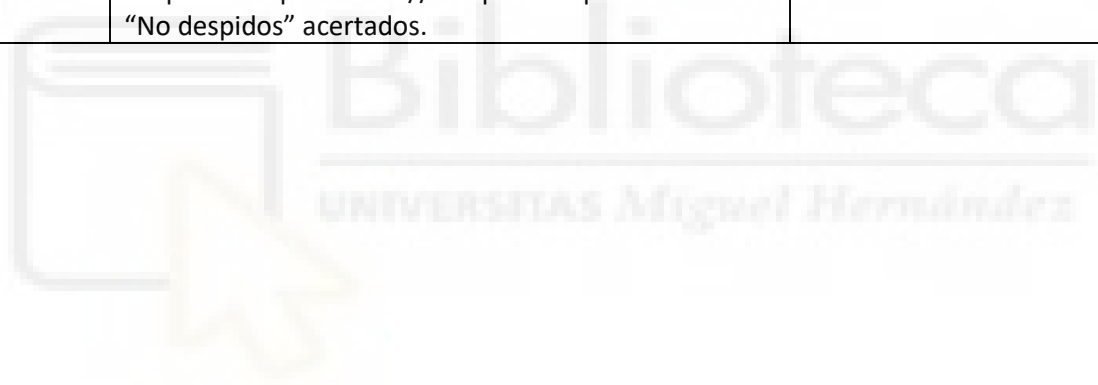
Autor/es	Objetivo	Muestra	Métodos	Resultados	Conclusiones
(Fernandez -Navarro et al., 2019)	- Evaluar la efectividad de los estilos de juego en fútbol, e influencia de variables contextuales (marcador, lugar y calidad oponente) sobre estos.	<p><u>Datos provistos por 'STATS LLC':</u> <u>Estilos juego:</u> 68.766 posesiones de 380 partidos Premier League 2015-2016. <u>Modelo xG:</u> 31.384 disparos Premier League 2011-2014</p>	<ul style="list-style-type: none"> - Desarrollan “Índice de efectividad de posesión” (PEI) sumando valores BMP y xG de la posesión cuando ésta acaba en tiro. - Modelo xG basado en distancia y tipo de disparo. - BMP indica peligro que cada movimiento de balón genera al rival. Cuantifica probabilidad de que cada pase de una posesión contribuya a un disparo posterior en la jugada (según datos pasados y valores xG). (Anexo 3) - Cuantificaron pertenencia porcentual de cada posesión a 8 estilos de juego (Anexo 4) - Multiplican PEI de cada posesión por su puntuación de pertenencia porcentual a cada estilo de juego → Obtienen puntuación efectividad de cada estilo de juego en cada posesión. 	- Ver efectividad de cada estilo de juego bajo la influencia de variables contextuales en Anexo 5.	<ul style="list-style-type: none"> - Índice de efectividad de posesión (PEI), calculado con métricas xG y BMP, puede ser usado como medida para cuantificar efectividad de las posesiones de equipo. - Efectividad de los estilos de juego analizados fue diferente según influían los factores contextuales. - Evaluar efectividad estilos de juego equipo propio o rival bajo circunstancias específicas útil para: apoyar decisiones estratégico-tácticas, mejorar preparación de partido y diseñar tareas de entrenamiento específicas.

		<u>Puntos movimiento de balón (BMP):</u> Posesiones Premier League 2009-2015.	- Aplican regresión estadística (modelo lineal mixto) para estimar relación entre efectividad estilos de juego y variables contextuales (marcador, lugar y calidad oponente).		
(Brîndescu et al., 2021)	- Analizar eficiencia presión alta Premier League 20/21.	Datos estadísticos (fbref.com) y partidos Premier League temporada 20/21 hasta el 20 de abril.	- Análisis datos estadísticos (fbref.com): % éxito (conseguir posesión en menos de 5s aplicando presión) del nº de veces que equipo aplicaba presión, tercio del campo donde se aplicaba la presión, zonas del campo donde equipos tocan más el balón, xG totales equipo y puesto clasificación. - Visionado partidos: determinar altura bloque defensivo (alto, medio, bajo) equipos analizados. - Analizan eficiencia de disparo (diferencia goles reales marcados y xG) de los equipos.	<u>Equipos que presionan tercio atacante del campo:</u> - En general, mayor % éxito en la presión. - Tocan más balones en el tercio atacante y área de penalti rival. - Mayor cantidad xG total equipo.	- Equipos con mayor % éxito en presión alta, entran en el tercio atacante más veces que los que no la aplican → esto permitiría crear más ocasiones y/o de mayor peligro dada la proximidad a la portería rival. - Generar ocasiones tras recuperar posesión o materializarlas depende de otros factores.

3.3 MITIGACIÓN DE SESGOS DE RESULTADO EN EL ANÁLISIS DE RENDIMIENTO

Autor/es	Objetivo	Muestra	Métodos	Resultados	Conclusiones
(Brechot & Flepp, 2020)	- Presentar método de evaluación de rendimiento en fútbol orientado al proceso en lugar de al resultado.	170.688 tiros en 7.304 partidos de La Liga, Premier League, Bundesliga, Ligue 1 y Serie A 13/14 – 16/17.	- Calculan modelo xG mediante regresión logística en base a distancia, tipo de jugada y parte del cuerpo con la que remata. - Regresión lineal puntos ganados previamente y puntos futuros; y ratio xG creados y concedidos, y puntos futuros. Comparan R2 de ambos casos. - Crean gráfico comparativo entre ranking oficial Premier League y ranking en base a ratio xG creados y concedidos; a mitad de temporada. - Mismo gráfico para muestra completa.	- R2 “ratio xG” mayor que R2 “puntos ganados previamente”; respecto a “puntos futuros”. - Diferencias entre valores R2 no significativa a partir de datos de 20 partidos. - 15 de 20 equipos Premier diferencia ≤ 3 puestos entre ambos rankings. - Observan casos de equipos sobre/infra recompensados en el ranking oficial de la liga.	- Puntos futuros estimados mejor por ratio xG que por puntos pasados. - Varianza puntos futuros explicada mejor por ratio xG que por puntos pasados. - A mayor muestra menor componente aleatorio. - Gran diferencia de posición en ambos rankings puede indicar influencia sustancial de elementos aleatorios. - Gráfica útil para tomar decisiones orientadas al proceso y no al resultado, reduciendo la influencia del componente aleatorio.

<p>(Flepp & Franck, 2021)</p>	<ul style="list-style-type: none"> - Clasificar despidos entrenadores de fútbol profesional en acertados e imprudentes. - Evaluar efecto sobre rendimiento de cada tipo de despido. 	<ul style="list-style-type: none"> - 9.130 partidos La Liga, Premier League, Bundesliga, Ligue 1 y Serie A 13/14 – 17/18. - 143 despidos involuntarios. - 214.194 tiros. 	<ul style="list-style-type: none"> - Cálculo modelo xG → regresión logística en base a distancia, tipo de jugada y parte del cuerpo con la que remata, efectos fijos temporada equipos para mitigar sesgos cualidades equipos marcar/conceder goles. - Comparan rankings oficial liga (OLT) y en base a ratio xG. - Despido acertado = ranking ratio xG igual o peor OLT = Grupo control (GC) “No despido” imprudente. - Despido imprudente = ranking ratio xG mejor que OLT = GC “No despido” acertado. - Sorpresa acumulada= diferencia puntos esperados y puntos reales momento del despido. - Comparan rendimiento posterior con GC “sorpresa acumulada” similar: Despidos acertados ↔ “No despidos” imprudentes // Despidos imprudentes ↔ “No despidos” acertados. 	<ul style="list-style-type: none"> - Rendimiento aumenta tras destitución acertada, no tras imprudente. - Aumenta igualmente tras destitución imprudente que tras “no destitución” acertada. 	<ul style="list-style-type: none"> - Variación rendimiento post-destitución depende de rendimiento pre-despido. - Reemplazar entrenador es beneficioso precedido de mal rendimiento en campo. No efecto si despido viene dado por mala suerte en resultados. - Destituciones imprudentes no afectan rendimiento deportivo → graves consecuencias financieras. - Importancia enfoques analíticos en la evaluación de rendimiento → reducir sesgo resultados.
-----------------------------------	---	---	---	--	--



4. DISCUSIÓN

En cuanto al **desarrollo del “modelo de goles esperados (xG)”**, Brechot & Flepp (2020), Flepp & Franck (2021) y Rathke (2017) sólo tuvieron en cuenta los datos de evento de disparo/remate; mientras que Anzer & Bauer (2021) además, incluyeron los datos de posicionamiento de todos los jugadores, árbitro y balón durante el partido. Esto permitió contemplar otros factores determinantes en la estimación de la probabilidad de éxito de cada disparo, como la presión defensiva sobre el jugador que chuta o la distancia entre el portero y la portería en el momento de disparo. Para ello, desarrollaron un algoritmo capaz de sincronizar los datos de posicionamiento y evento de manera automática dado el gran tamaño de la muestra, consiguiendo además que su modelo tenga una aplicación real en el día a día de la práctica profesional. Este algoritmo sufría un desfase absoluto de sólo $0,23 (\pm 0,49)$ s comparado con el “gold standard” de sincronización aplicado para ambas bases de datos; suponiendo una enorme mejora en términos de agilidad respecto al método manual, por lo que los autores sugieren que el desarrollo de algoritmos similares adaptados a otras disciplinas deportivas o al estudio de otros datos de evento (ej.: pases y tackles) puede presentar un progreso sustancial para el análisis del rendimiento deportivo.

En cuanto al cálculo de la probabilidad de gol de cada chut, Rathke (2017) lo realizó manualmente, distribuyendo cada disparo en 8 zonas diferentes del campo y obteniendo la probabilidad de éxito (valor xG) en cada una de ellas, teniendo en cuenta indirectamente distancia y ángulo a portería. Por el contrario, Brechot & Flepp (2020) aplicaron un análisis de regresión logística sobre los tiros de la muestra en base a distancia, tipo de jugada y parte del cuerpo con la que se remata. En la misma línea, Flepp & Franck (2021) añadieron a este mismo modelo los “efectos fijos” de las cualidades de los equipos para marcar y conceder goles, mitigando posibles sesgos derivados de no contemplar dichas individualidades. Por otro lado, Anzer & Bauer (2021) entrenaron y compararon 5 modelos diferentes de aprendizaje automático supervisado, resultando como más preciso un modelo de “potenciación de gradiente” (regresor y clasificador estadístico) que combinaba otros tres submodelos diferenciados del mismo tipo en función de si el disparo era un chut con la pierna, un remate de cabeza o un libre directo. En este último, las características influyentes en la ocasión de gol (Anexo 2) fueron determinadas junto con analistas tácticos profesionales. El análisis posterior indicó que los factores más determinantes en la probabilidad de marcar eran la localización del disparo (distancia y ángulo a portería) y la distancia del portero a la portería. Cabe destacar, que ninguno de los autores contempló los goles en propia puerta es sus estudios puesto que son de carácter involuntario y, por lo tanto, su inclusión podría alterar el cálculo de probabilidades.

El modelo propuesto por Anzer & Bauer (2021) parece ser el más completo de entre todos los revisados, con un error promedio de estimación de goles esperados respecto a la muestra total de 0.0928 goles. Probablemente, tal y como reflejan los autores, esto sea debido a la combinación de contar con una muestra mayor que en otros artículos, la inclusión de datos posicionales gracias a su algoritmo de sincronización, y a todos los factores concernientes al disparo tenidos en cuenta para el cálculo probabilístico. Además, es el único estudio en el que se prueba el poder predictivo del modelo sobre datos no vistos. No obstante, enfoques más simples como los de Brechot & Flepp (2020), Flepp & Franck (2021), o Rathke (2017) que también presentaba valores elevados entre goles reales marcados y xG en la prueba estadística de bondad del ajuste del modelo (R^2), pueden ser interesantes para su aplicación en contextos con menor disponibilidad de recursos o en los que sólo contamos con datos de evento y no posicionales.

Respecto a la **aplicación del modelo xG para el análisis de rendimiento en fútbol**; Anzer & Bauer (2021) y Rathke (2017) compararon el número real de goles marcados en relación con las métricas xG, con el objetivo de estudiar la eficiencia de tiro de los equipos y/o jugadores. Las probabilidades de gol son determinadas en base a los valores promedio de la muestra, por lo

que las estimaciones no contemplan la habilidad individual del rematador ni del portero en cada caso. Esto supone una limitación para las tareas de predicción; ya que el modelo generalmente subestima los xG de los equipos mejor clasificados que suelen contar con jugadores más hábiles y con mayor eficiencia goleadora, capaces de materializar ocasiones de menor calidad, y sobreestima los xG de los equipos peor clasificados, siguiendo el mismo razonamiento para jugadores menos hábiles. No obstante, esta limitación del modelo permite identificar jugadores que, dada su habilidad individual, rinden consistentemente por encima, por debajo o próximos a sus xG en el largo plazo (Anzer & Bauer, 2021; Rathke, 2017). Gracias a los xG es posible obtener una evaluación más objetiva de la eficiencia goleadora de los equipos/jugadores que la realizada simplemente en base al número de tiros o a los propios goles, ya que incluyen en el análisis la calidad de las ocasiones de las que han dispuesto. Además, Anzer & Bauer (2021) proponen que los valores xG pueden extenderse como “asistencias esperadas (xA)”; es decir asignar el valor xG al asistente, determinando la probabilidad de que un pase previo a un tiro se materialice como asistencia. Esto permite evaluar las cualidades asistentes de un jugador independientemente de la habilidad del rematador para marcar. El modelo xG puede ser una herramienta clave para el scouting; especialmente identificando jugadores que rinden consistentemente por encima de sus xG o que destacan en cuanto a valores xA, o a porteros que conceden menos goles que sus xG en contra.

En línea con lo que sugiere Rathke (2017), el modelo xG también puede emplearse para mejorar el análisis táctico y el entendimiento del juego. Anzer & Bauer (2021) estudiaron el comportamiento de disparo de equipos/jugadores comparando el número de tiros por partido con el valor xG promedio por disparo; analizando si solían chutar un mayor número de veces a puerta, pero con menor peligro, o si por el contrario trataban de disparar sólo en ocasiones muy claras de gol, etc. Este análisis permite adaptar la estrategia defensiva en función del comportamiento de disparo rival.

Por su parte, Brîndescu et al. (2021) en su estudio sobre la eficiencia de la presión alta; usaron los xG para cuantificar la calidad de las ocasiones creadas en comparación con otros estilos de juego, y para explicar posibles causas por las que algunos equipos que aplicaron la presión alta con efectividad no consiguieron buenos resultados en comparación con los que no (ej.: pobre eficiencia de tiro), mitigando el componente resultadista de los goles en el análisis del estilo de juego.

Continuando con el estudio sobre el planteamiento estratégico, Fernández-Navarro et al. (2019) desarrolló un “Índice de efectividad de posesión” (PEI), que calculó sumando los valores BMP (Puntos de movimiento de balón) al valor xG de cada posesión si esta acababa en disparo, para evaluar la efectividad de ocho estilos de juego. Los BMP indican el peligro que cada movimiento de balón genera al rival, cuantificando la probabilidad de que cada pase de una posesión contribuya a un disparo posterior en la jugada, según datos pasados y valores xG (Anexo 3).

Además, gracias a la sincronización de datos posicionales en el modelo de Anzer & Bauer (2021), se abre la posibilidad de analizar cómo influyen estos sobre los datos de evento; por ejemplo, estudiando cómo fluctúa el valor xG hipotético de un tiro antes y después de un regate o un pase, cuantificando indirectamente la contribución de esa acción a la jugada ofensiva, y apoyando la evaluación sobre la toma de decisiones del jugador.

Otra utilidad de esta sincronización es estudiar la influencia de las características concernientes al disparo sobre la probabilidad de marcar; por lo que entrenadores y jugadores pueden analizar objetivamente qué tipo de situaciones de tiro correlacionan mejor con altas probabilidades de gol, orientando la estrategia ofensiva y las tareas de entrenamiento hacia tratar de maximizar la cantidad de veces que alcanzan ese tipo de situaciones. El mismo razonamiento puede emplearse a nivel defensivo, tratando de evitar que el equipo rival alcance situaciones con altas probabilidades de gol.

Todos estos enfoques reflejan las distintas aplicaciones que tienen los xG para el análisis estratégico-táctico; examinando la información que ofrece la métrica en sí misma, comparándola con otros datos, o incluso incluyéndola en la computación de otros modelos más complejos para estudiar la producción ofensiva de jugadores y equipos en distintas situaciones.

En cuanto a la **influencia del azar en el análisis de resultados y la toma de decisiones**; Brechot & Flepp (2020) demostraron el valor informativo que ofrece la ratio xG respecto a los puntos pasados, sobre los puntos futuros; explicando un mayor porcentaje de la varianza en los resultados. Para ello, tras realizar un análisis de regresión lineal por separado de ambas variables sobre los puntos futuros, compararon el R²; siendo mayor el de la ratio xG, especialmente cuanto mayor era la muestra de datos previos. Esto permite tomar decisiones orientadas al proceso, analizando con mayor objetividad el rendimiento real en el campo y mitigando la influencia de los factores aleatorios en el análisis de resultados.

Brechot & Flepp (2020) y Flepp & Franck (2021) crearon un gráfico comparativo entre la clasificación oficial de la liga y un ranking en base a la ratio xG marcados y concedidos. Según lo explicado anteriormente, una gran diferencia de posición en ambos rankings podría indicar una influencia sustancial de los factores aleatorios sobre los resultados. Flepp & Franck (2021) emplearon este método para clasificar las destituciones o no destituciones de entrenadores en acertadas o imprudentes. Este enfoque analítico permite justificar decisiones frente a aficionados y medios, mitigar sesgos de resultado, y de este modo mejorar el rendimiento económico y deportivo del club. Por su parte, Anzer & Bauer (2021), identificaron que, obviando los empates, en el 73,3% de los partidos el equipo ganador consiguió un mayor valor xG agregado, mientras que en sólo un 56,2% obtuvo un mayor número de tiros totales. Siguiendo este razonamiento compararon los valores xG agregados entre equipos, diferenciando entre victorias “merecidas” y “no merecidas”. Un mayor valor xG agregado implicaba que habían generado un mayor número de ocasiones y/o con mayor probabilidad de marcar.

Por último, en cuanto a la **capacidad predictiva del modelo xG** Anzer & Bauer (2021) y Brechot & Flepp (2020) presentaron el potencial de la ratio xG marcados y concedidos para predecir la ratio de goles y los puntos futuros respectivamente, en comparación a otras métricas estudiadas por separado. En ambos casos, las estimaciones fueron más precisas cuanto mayor era la muestra de datos previos recientes.

Esto abre nuevas posibilidades de investigación para el análisis de rendimiento en fútbol sobre las utilidades que ofrece la capacidad de predecir resultados futuros en base a datos actuales.

4.1 LIMITACIONES

Esta revisión presenta varias limitaciones a tener en cuenta. En primer lugar, aunque todos los modelos xG presentados se fundamentan en la misma idea, son desarrollados siguiendo metodologías distintas y no presentan un método de evaluación estandarizado que permita compararlos directamente. Además, la muestra es muy dispar entre algunos estudios, y los datos analizados no son extrapolables directamente a otros contextos. Por último, sólo un estudio utiliza las estimaciones de su modelo sobre datos no vistos, siendo necesarios más trabajos orientados en esta línea (Anzer & Bauer, 2021).

4.2 CONCLUSIONES

Combinar datos posicionales y de evento en el desarrollo de un modelo xG permite contemplar y estudiar más características relativas al disparo determinantes en la consecución del éxito, y por lo tanto en la estimación de las probabilidades de gol.

El análisis de la información que ofrece la métrica xG en sí misma o junto a otros datos hace posible llevar a cabo estudios sobre el comportamiento y la eficiencia de disparo, el planteamiento estratégico-táctico, o evaluar el rendimiento de un asistente (xA) independientemente de la ejecución final del rematador.

La ratio xG marcados y concedidos permite predecir resultados futuros con mayor precisión que otras métricas tradicionales.

Además, gracias al modelo xG es posible filtrar los factores aleatorios que influyen sobre los resultados, orientando la evaluación del rendimiento y la consecuente toma de decisiones al proceso en lugar de a los resultados.

5. PROPUESTA DE INTERVENCIÓN

Teniendo en cuenta los resultados obtenidos en esta revisión, el objetivo de la siguiente propuesta de intervención es utilizar la información ofrecida por un modelo de goles esperados (xG) para apoyar las decisiones de la dirección deportiva en un club de fútbol profesional de la Primera División de España.

Se sitúa en la jornada 16 de liga y todos los clubes disponen de los datos xG actualizados de todos los equipos y jugadores en competiciones profesionales de la temporada pasada y actual, ofrecidos por el proveedor de datos oficial de LaLiga (StatsPerform).

En cuanto al mercado de fichajes de invierno, el director deportivo tiene como objetivos prioritarios encontrar un delantero centro sub-23 que se adapte a la filosofía de juego del club y a una media punta de corte creativo que genere claras ocasiones de gol a partir del último pase. El presupuesto del club es limitado, por lo que quedan fuera de su alcance fichajes estrella con gran impacto mediático. Además, en los últimos 5 partidos el equipo no ha obtenido buenos resultados y tanto la prensa local como la gran mayoría de los aficionados parecen estar mostrando su descontento con la plantilla, presionando a la junta directiva para una posible destitución del entrenador a mitad de temporada. La junta ha encomendado personalmente al director deportivo un informe de evaluación del rendimiento del equipo para tomar las decisiones oportunas antes del cierre del mercado invernal.

Para solucionar estos problemas y cumplir con los objetivos asignados, se propone a la dirección deportiva aplicar un enfoque analítico fundamentado en el estudio de las métricas xG.

En cuanto a las tareas de **identificación e incorporación de talento (scouting)**, se comparan el número total de goles con los valores xG agregados de todos los delanteros centro sub-23 disponibles, normalizados por 90 minutos. Además, puesto que debe adaptarse a la filosofía de juego del club, se realizará un estudio del comportamiento de disparo del equipo; comparando el número de disparos por partido del equipo y el valor xG promedio de cada ocasión, concluyendo que en general el equipo prefiere no arriesgar la posesión del balón y chutar sólo cuando encuentra ocasiones claras de gol. Esta misma comparativa se aplicará a los delanteros de la muestra para identificar, según los datos, cuáles muestran un comportamiento de disparo similar, y potencialmente serán hábiles a la hora de realizar movimientos en el área rival que les permitan generar ocasiones más peligrosas. Por otro lado, los valores xG serán trasladados a métricas de asistencias esperadas (xA) en los centrocampistas para determinar la probabilidad de que sus pases acaben materializándose como asistencia independientemente de la ejecución final del rematador, y para estudiar el comportamiento en el último pase de los jugadores, tratando de encontrar aquellos que se adapten al comportamiento de disparo del equipo y mantengan el control del balón hasta que encuentren ocasiones claras de gol; es decir, que den un menor número de últimos pases por partido pero que estos tengan un valor promedio xA elevado.

Para ajustarse al presupuesto del club, dentro de los jugadores que potencialmente se adaptan al comportamiento de disparo del equipo, se pretende encontrar delanteros y centrocampistas infravalorados por el mercado. El objetivo serían delanteros que, en el equivalente de una muestra de al menos 15 partidos de 90 minutos, rindan consistentemente por encima de sus xG aunque el número total de goles no sea muy elevado, por lo que a pesar de haber dispuesto de menos ocasiones o de menor peligro, su habilidad individual les ha permitido marcar más goles de los esperados. Además, se buscarán centrocampistas con valores xA muy elevados en la misma muestra en comparación con las asistencias reales, lo que significará que, a pesar de haber ofrecido ocasiones claras de gol gracias a su último pase, la ejecución final del rematador no ha sido buena. En ambos casos, las métricas tradicionales de goles y asistencias no reflejan el rendimiento potencial de los jugadores identificados, por lo que es más probable que su valoración en el mercado sea menor a la esperada.

En cuanto a la **evaluación de los resultados del equipo**; se creará un gráfico comparativo entre la clasificación oficial de la liga y un ranking basado en la ratio xG marcados y concedidos (Brechot & Flepp, 2020). En este caso, aunque en el ranking xG el equipo se encuentra en los puestos esperados al principio de temporada, en el oficial ocupa 5 posiciones por debajo. Este análisis indica que es posible que el azar esté influyendo negativamente de manera sustancial en los resultados, por lo que según Flepp & Franck (2021) se podría calificar la decisión de destituir al entrenador de imprudente, no obteniendo mejoras probablemente en el rendimiento futuro en comparación con mantener al entrenador, y acarreando graves consecuencias financieras para el club. Además, la junta directiva podrá mostrar datos para justificar sus decisiones frente a los aficionados y medios de comunicación, aliviando la presión ejercida por estos.

6. BIBLIOGRAFÍA

- Anzer, G., & Bauer, P. (2021). A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Frontiers in Sports and Active Living*, 3. <https://doi.org/10.3389/fspor.2021.624475>
- Brechot, M., & Flepp, R. (2020). Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. *Journal of Sports Economics*, 21(4), 335–362. <https://doi.org/10.1177/1527002519897962>
- Brîndescu, S., Datcu, F-R., & Buda, I-A. (2021). Study on the efficiency of advanced pressing in the Premier League. *Baltic Journal of Health and Physical Activity*, 115–122. <https://doi.org/10.29359/BJHPA.13.Spec.Iss1.11>
- Eaves, S. J. (2015). A history of sports notational analysis: A journey into the nineteenth century. *International Journal of Performance Analysis in Sport*, 15(3), 1160–1176. <https://doi.org/10.1080/24748668.2015.11868859>
- Fernandez-Navarro, J., Fradua, L., Zubillaga, A., & McRobert, A. P. (2019). Evaluating the effectiveness of styles of play in elite soccer. *International Journal of Sports Science and Coaching*, 14(4), 514–527. <https://doi.org/10.1177/1747954119855361>
- Flepp, R., & Franck, E. (2021). THE PERFORMANCE EFFECTS OF WISE AND UNWISE MANAGERIAL DISMISSALS. *Economic Inquiry*, 59(1), 186–198. <https://doi.org/10.1111/ecin.12924>
- Gauriot, R., & Page, L. (2019). Fooled by Performance Randomness: Overrewarding Luck. *The Review of Economics and Statistics*, 101(4), 658–666. https://doi.org/10.1162/rest_a_00783

- Lago-Peñas, C., Lago-Ballesteros, J., Dellal, A., & Gómez, M. (2010). Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league. In *Journal of Sports Science and Medicine* (Vol. 9). <http://www.jssm.org>
- Lago-Peñas, C., Lago-Ballesteros, J., & Rey, E. (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, 27(1), 135–146. <https://doi.org/10.2478/v10078-011-0011-3>
- Lefgren, L., Platt, B., & Price, J. (2015). Sticking with what (Barely) worked: A test of outcome bias. *Management Science*, 61(5), 1121–1136. <https://doi.org/10.1287/mnsc.2014.1966>
- Memmert, D., Hüttermann, S., Hagemann, N., Loffing, F., & Strauss, B. (2013). Dueling in the penalty box: Evidence-based recommendations on how shooters and goalkeepers can win penalty shootouts in soccer. *International Review of Sport and Exercise Psychology*, 6(1), 209–229. <https://doi.org/10.1080/1750984X.2013.811533>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., Estarli, M., Barrera, E. S. A., Martínez-Rodríguez, R., Baladia, E., Agüero, S. D., Camacho, S., Buhning, K., Herrero-López, A., Gil-González, D. M., Altman, D. G., Booth, A., ... Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1). <https://doi.org/10.1186/2046-4053-4-1>
- Morgulev, E., Azar, O. H., & Lidor, R. (2018). Sports analytics and the big-data era. *International Journal of Data Science and Analytics*, 5(4), 213–222. <https://doi.org/10.1007/s41060-017-0093-7>
- Pollard, R. & Reep, C. (1997). *Measuring the effectiveness of playing strategies at soccer*. J. R. Stat. Soc. D Stat. 46, 541–550. doi: 10.1111/1467-9884.00108
- Rathke, A. (2017). An examination of expected goals and shot efficiency in soccer. *Journal of Human Sport and Exercise*, 12(Proc2). <https://doi.org/10.14198/jhse.2017.12.proc2.05>
- Szymanski, S. (2020). Sport analytics: Science or alchemy? *Kinesiology Review*, 9(1), 57–63. <https://doi.org/10.1123/kr.2019-0066>
- Tenga, A., Ronglan, L. T., & Bahr, R. (2010). Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science*, 10(4), 269–277. <https://doi.org/10.1080/17461390903515170>

7. ANEXOS

Anexo 1:

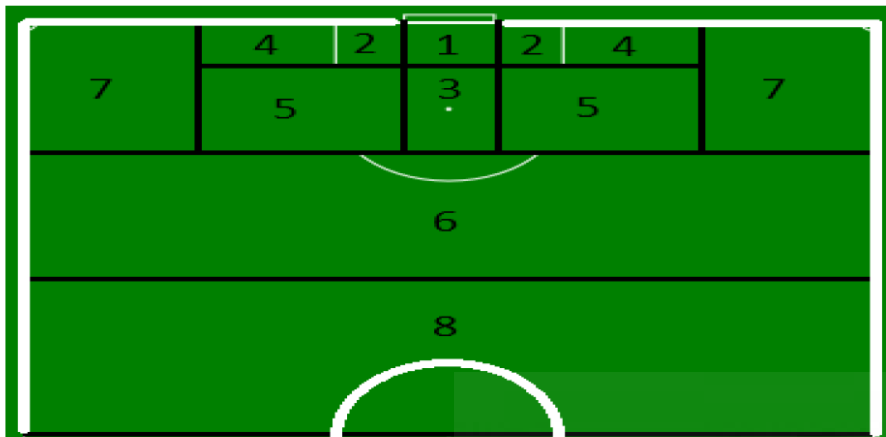


Figure 1. Pitch zones.

Anexo 2:

TABLE 1 | Features derived from synchronized positional and event data used to train our model.

Feature	Value	Description
Shot location	Numeric	The x, y and the z-coordinate of the ball at the time of the shot are used for several features, such as angle and distance to goal center.
Speed of player taking the shot	Numeric	The speed of the player attempting the shot, at the time of the shot (in <i>km/h</i>).
Defenders in the line of the shot	Numeric	The number of defenders in the line of the shot.
Goalkeeper position	Numeric	The position of the goalkeeper is used for two different features, describing whether they are in the line of shot and their distance to the goal.
Pressure on the player taking the shot	Numeric	Various metrics describing the pressure that the player was under while attempting the shot, at the time of the shot Andrienko et al., 2017.
Type of shot	Categorical	Describing the body part used for the shot (<i>Head, leg or other</i>).
Taker ball-control	Categorical	Describes how the player taking the shot gained control of the ball before/when taking the shot (<i>volley, controlShot, dribblingLess10m, dribblingMore10m, setPiece</i>).
After freekick	Categorical	Indicates whether the shot followed a freekick.
Freekick	Categorical	Describes whether the shot is a direct freekick or not.

Anexo 3:

Table 1. Styles of play definitions by STATS LLC.

Style of play	Definition
Direct play	Captures instances of play where teams attempt to move the ball quickly towards the opposition's goal through the use of long passes. Specifically, it looks at the distance gained forward every time a team makes use of any of the following events: pass, direct free-kick pass, indirect free-kick pass, cross, direct free-kick cross, indirect free-kick cross, goal kick, goalkeeper throw, goalkeeper kick, throw in, or clearance. The forward distance gained must be greater than 20m and reaches 100% at 40m
Counterattack	A team regains possession and moves the ball into an attacking area via passes, dribbles or a combination of both. The ball must reach a target location within the opposition's half. This location varies depending on the regain location. The speed of the transition from a regain to a target location determines the Counterattack value. The quicker the ball is moved up the pitch, the higher the Counterattack value. Counterattack regains include: goal keeper catch, goal keeper save, interception, clearance, header, tackle and block. Counterattack distance gained include: touch, dribbling, clearance and pass.
Maintenance	Captures possessions in which a team looks to maintain possession of the ball within the defensive area of the pitch. The time spent in possession directly relates to the Maintenance membership value. The team must have a passage of play lasting more than 10 s. From then on, the membership value increases linearly up until 30s where it reaches 100%.
Build up	Captures long and controlled ball possessions – but is aimed at periods of play where a team is looking for opportunities to attack. The calculation is similar to Maintenance with the differences being the zone on the pitch and the time thresholds. The Build Up area is between the halfway line and the opposition's penalty area and the passage of play must last more than 8 s. From then on, the membership value increases linearly up until 25s where it reaches 100%.
Sustained threat	Similar to Maintenance and Build Up. However, here the focus lies on possessions in the attacking third of the pitch. The time spent in possession must be more than 6 s, reaching 100% at 20 s.
Fast tempo	Captures when the team is moving the ball quickly to increase the tempo and speed of the game. Fast Tempo looks at sequences of consecutive individual 'fast possessions'. An individual fast possession must occur in the opposition's half and can be achieved as follows: the player releases the ball to a team mate in less than 2 s, or the player dribbles at a high tempo.
Crossing	It occurs if the ball is delivered from a wide area of the pitch with the intention of finding a teammate. All Crossing events in a possession are assigned a value of 100%. The value assigned to the team possession can only be 0% or 100% depending on the occurrence of a crossing event. Crossing events are: cross, corner cross, direct free-kick cross and indirect free-kick cross.
High pressure	Captures how high up the pitch teams regain possession. The first factor taken into consideration is the location where the team wins the ball: High Press regains are those higher than 5 m prior to the halfway line. The value increases linearly up until 15m into the opposition's half where it reaches 100%. The second factor is the opposition's time in possession prior to the High Press regain happening. To retain the full value established based on the regain location, the opposition must have been in possession for at least 10s. This time factor is introduced to try and capture controlled pressing efforts rather than 'counter press' regains. The combination of these two factors leads to the final High Press membership value. Regain events include: interception, header, tackle and block.

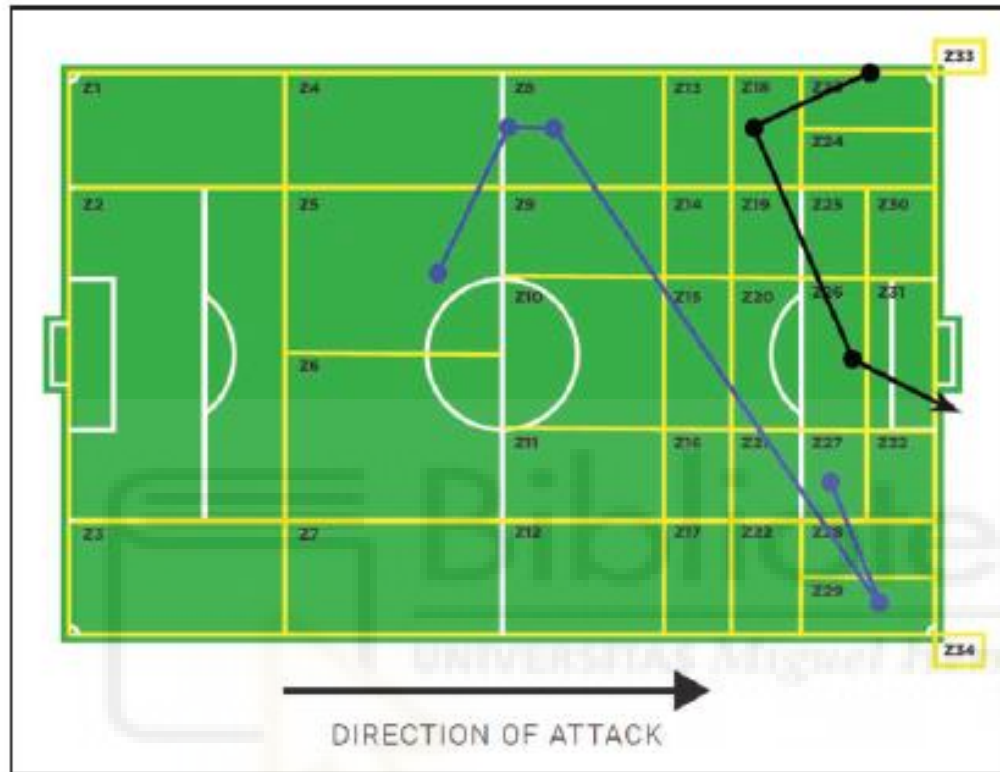


Figure 3. Pitch division to calculate ball movement points (BMP). Black chain of ball moves ending in a header shows a total of 0.15 PEI ($0.004 + 0.032$) BMP + 0.114 xG. Blue chain of ball moves ending in a possession lost shows a total of -0.009 PEI ($0.001 + 0.01 + 0.011 - 0.031$) BMP.

Anexo 5:

Table 2. Effectiveness of the eight styles of play controlling for contextual variables.

Fixed effects	Direct Play				Counterattack			
	β (SE)	95% CI	t	P	β (SE)	95% CI	t	P
Intercept	0.648 (0.035)	0.579, 0.717	18.342	<0.001	1.508 (0.078)	1.356, 1.660	19.448	<0.001
Match status (-2 or more)	0.178 (0.062)	0.056, 0.301	2.856	0.004	-0.040 (0.184)	-0.400, 0.320	-0.218	0.828
Match status (-1)	0.038 (0.046)	-0.051, 0.128	0.840	0.401	-0.312 (0.136)	-0.580, -0.045	-2.289	0.022
Match status (+1)	-0.024 (0.045)	-0.111, 0.064	-0.532	0.595	0.355 (0.134)	0.093, 0.617	2.656	0.008
Match status (+2 or more)	0.155 (0.061)	0.036, 0.274	2.558	0.011	0.766 (0.171)	0.431, 1.100	4.482	<0.001
Venue (away)	-0.084 (0.032)	-0.146, -0.022	-2.657	0.008	-	-	-	-
Quality opposition	-0.018 (0.002)	-0.022, -0.013	-7.294	<0.001	-0.015 (0.006)	-0.028, -0.002	-2.313	0.023
Venue (away) \times Quality opposition	-	-	-	-	-	-	-	-
Random effects	Estimate	SD			Estimate	SD		
Match	0.017	0.132			0.097	0.311		
Venue	-	-			-	-		
Quality opposition	-	-			-	-		
Team	0.009	0.093			0.020	0.142		
Residuals	10.298	3.209			18.706	4.325		
$R^2_{(m)}$	0.002				0.007			
$R^2_{(c)}$	0.005				0.013			

(continued)

Table 2. Continued.

Fixed effects	Maintenance				Build Up			
	β (SE)	95% CI	t	P	β (SE)	95% CI	t	P
Intercept	0.660 (0.044)	0.573, 0.746	14.986	<0.001	1.496 (0.090)	0.335, 0.518	16.632	<0.001
Match status (-2 or more)	0.056 (0.061)	-0.063, 0.176	0.924	0.355	-	-	-	-
Match status (-1)	-0.009 (0.047)	-0.101, 0.082	-0.202	0.840	-	-	-	-
Match status (+1)	-0.060 (0.050)	-0.158, 0.038	-1.196	0.232	-	-	-	-
Match status (+2 or more)	0.151 (0.064)	0.025, 0.277	2.343	0.019	-	-	-	-
Venue (away)	-0.078 (0.033)	-0.143, -0.013	-2.350	0.019	-	-	-	-
Quality opposition	-0.015 (0.003)	-0.020, -0.009	-5.350	<0.001	-0.025 (0.005)	-0.017, -0.010	-4.904	<0.001
Venue (away) \times Quality opposition	-	-	-	-	-	-	-	-
Random effects	Estimate	SD			Estimate	SD		
Match	0.014	0.117			0.074	0.273		
Venue	-	-			-	-		
Quality opposition	-	-			-	-		
Team	0.022	0.147			0.139	0.373		
Residuals	6.683	2.585			16.272	4.034		
$R^2_{(m)}$	0.003				0.003			
$R^2_{(c)}$	0.008				0.016			

(continued)

Table 2. Continued.

Fixed effects	Sustained Threat				Fast Tempo			
	β (SE)	95% CI	<i>t</i>	<i>P</i>	β (SE)	95% CI	<i>t</i>	<i>P</i>
Intercept	2.153 (0.100)	0.425, 0.630	21.451	<0.001	2.872 (0.109)	2.659, 3.086	26.345	<0.001
Match status (-2 or more)	-	-	-	-	-	-	-	-
Match status (-1)	-	-	-	-	-	-	-	-
Match status (+1)	-	-	-	-	-	-	-	-
Match status (+2 or more)	-	-	-	-	-	-	-	-
Venue (away)	-	-	-	-	-	-	-	-
Quality opposition	-0.029 (0.007)	-0.016, -0.009	-4.012	<0.001	-0.032 (0.013)	-0.057, -0.007	-2.549	0.012
Venue (away) \times Quality opposition	-	-	-	-	-	-	-	-
Random effects	Estimate	SD			Estimate	SD		
Match	0.115	0.339			0.336	0.580		
Venue	-	-			-	-		
Quality opposition	-	-			0.006	0.081		
Team	0.155	0.394			0.073	0.269		
Residuals	29.240	5.407			31.682	5.629		
$R^2_{(m)}$	0.002				0.002			
$R^2_{(c)}$	0.011				0.035			

(continued)

Table 2. Continued.

Fixed effects	Crossing				High pressure			
	β (SE)	95% CI	<i>t</i>	<i>P</i>	β (SE)	95% CI	<i>t</i>	<i>P</i>
Intercept	5.053 (0.193)	4.675, 5.431	26.218	<0.001	0.678 (0.037)	0.605, 0.752	18.092	<0.001
Match status (-2 or more)	0.130 (0.367)	-0.590, 0.850	0.354	0.723	-	-	-	-
Match status (-1)	-0.473 (0.273)	-1.008, 0.062	-1.733	0.083	-	-	-	-
Match status (+1)	-0.026 (0.339)	-0.692, 0.639	-0.077	0.938	-	-	-	-
Match status (+2 or more)	1.446 (0.437)	0.589, 2.304	3.306	<0.001	-	-	-	-
Venue (away)	-	-	-	-	-0.119 (0.049)	-0.216, -0.023	-2.421	0.016
Quality opposition	-0.056 (0.015)	-0.086, -0.025	-3.600	<0.001	-0.023 (0.004)	-0.031, -0.014	-5.150	<0.001
Venue (away) \times Quality opposition	-	-	-	-	0.017 (0.006)	0.006, 0.029	2.891	0.004
Random effects	Estimate	SD			Estimate	SD		
Match	0.358	0.598			0.098	0.313		
Venue	-	-			0.067	0.259		
Quality opposition	-	-			<0.001	0.015		
Team	0.289	0.538			0.003	0.052		
Residuals	112.747	10.618			7.508	2.740		
$R^2_{(m)}$	0.004				0.003			
$R^2_{(c)}$	0.010				0.013			

β : beta coefficient; SE: standard error; CI: confidence interval. Statistical significance set at $P < 0.05$. Intercepts represent a draw and playing home.