

DATOS EN ABIERTO E INVESTIGACIÓN EN FUTBOL: DEL BOX SCORE AL GOOGLE TRENDS



El análisis deportivo, estadísticas o investigación deportiva ha cambiado mucho desde que comenzó en el siglo XIX hasta ahora, combinando datos, tecnología y estrategia. ¿Cómo influye la tecnología moderna en esto y qué ayuda nos brindan herramientas como Google Trends en la investigación actual sobre fútbol?

Antecedentes

El análisis deportivo (Sport analytics) se originó en 1858, cuando un periodista deportivo llamado Henry Chadwick desarrolló la primera herramienta estadística para medir cuantitativamente el rendimiento de los jugadores y equipos de béisbol: el box score. El box score (figura 1), comúnmente escrito en forma tabular, resume las acciones de juego de los jugadores durante un partido.

Figura 1: Una puntuación detallada de béisbol de 1876.

BOSTON.						ATHLETIC.							
T.	R.	lb.	PO.	A.	E.	T.	R.	lb.	PO.	A.	E.		
G. Wright, s.s.	6	4	4	1	5	2	Force, a. s.	5	1	2	1	3	2
Leonard, 2b.	6	3	3	4	4	3	Eggler, c. f.	5	3	3	0	0	0
O'Rourke, lb.	6	2	3	9	0	1	Fisler, r. f.	5	0	1	2	0	0
Murnan, l. i.	6	1	0	3	1	0	Meyerle, 3db.	5	1	2	2	3	3
Schafer, 2d b.	6	3	3	3	1	2	Sutton, 1st b.	5	1	2	10	0	0
McGinley, c. f.	6	0	0	0	0	1	Coons, c.	5	1	0	1	1	3
Manning, r. f.	6	0	2	2	0	0	Hall, l. f.	5	1	3	5	0	0
Morrill, c.	6	2	2	4	1	2	Fowler, 2d b.	6	1	2	6	7	5
Josephs, p.	5	4	4	1	1	2	Knight, p.	5	2	2	0	1	2
Totals	53	19	21	27	13	13	Totals	46	11	17	27	15	15
Boston	9	1	3	3	4	1	0	2	5	19			
Athletic	1	0	0	0	3	3	2	2	0	11			

Runs earned—Boston, 4; Athletic, 5. Home-run—Hall, 1.
 Total bases on hits—Boston, 22; Athletic, 20. First base by
 errors—Boston, 8; Athletic, 5. Umpire, George White of
 Lowell, Mass. Time 2h. 47m.

Fuente: Wikipedia.com

Años más tarde, en 1861, Chadwick escribió la primera guía documentada de análisis deportivo. Este libro defendía la necesidad de analizar el rendimiento de los jugadores en el campo para estimar sus habilidades. Sin embargo, no fue hasta principios del siglo XX cuando el análisis deportivo llamó la atención de analistas y empresas.

En este contexto, investigadores y analistas vieron la necesidad de implantar sistemas de información y modelos para evaluar los datos de rendimiento. Así, surgió la analítica deportiva como un proceso de búsqueda, interpretación y procesamiento de datos que proporcionan ventajas competitivas, ventajas competitivas que están relacionadas tanto con el ámbito deportivo como con el económico.

Más recientemente.

Con la aparición disruptiva de la Industria 4.0, la cual combina técnicas avanzadas de producción y operaciones con tecnologías inteligentes que se integran en las organizaciones, las personas y los activos, las posibilidades de ambos ámbitos se han incrementado enormemente. Esa evolución ha hecho que el análisis deportivo suponga un campo emergente que brinda a los investigadores la oportunidad de estudiar muchos problemas de utilidad para el deporte.

Las áreas de análisis, así como la finalidad de los estudios, difieren según el deporte y los datos disponibles. Por ejemplo, en el caso del fútbol, los equipos reciben grandes cantidades de datos, ya sea a través de proveedores externos o como resultado de la información relacionada con los jugadores que ellos mismos recopilan (análisis de sangre o dispositivos de sistemas de posicionamiento global (GPS)).

La categorización de la analítica futbolística según la naturaleza de los datos disponibles es en eventing y tracking.

Los datos de eventing (eventos) son las acciones de juego medibles, relacionadas con el balón, realizadas durante un partido de fútbol (por ejemplo, número de goles, número de asistencias o número de centros). Estos datos pueden obtenerse a través de proveedores de análisis deportivos (Opta, Wyscout o STATS), equipos de fútbol y sitios web abiertos dedicados al seguimiento de estadísticas como <https://fbref.com/es/>.

Los datos de tracking (seguimiento) recogen información sobre el movimiento de los jugadores (con y sin balón) sobre el terreno de juego, es decir, datos espaciotemporales que captan la posición de los jugadores en cada segundo del partido. A diferencia de los datos de eventos, los datos de seguimiento son más difíciles de encontrar ya que, en la mayoría de los casos, no están disponibles en sitios web públicos y sólo pueden obtenerse de empresas de datos o equipos de fútbol.

Desafíos

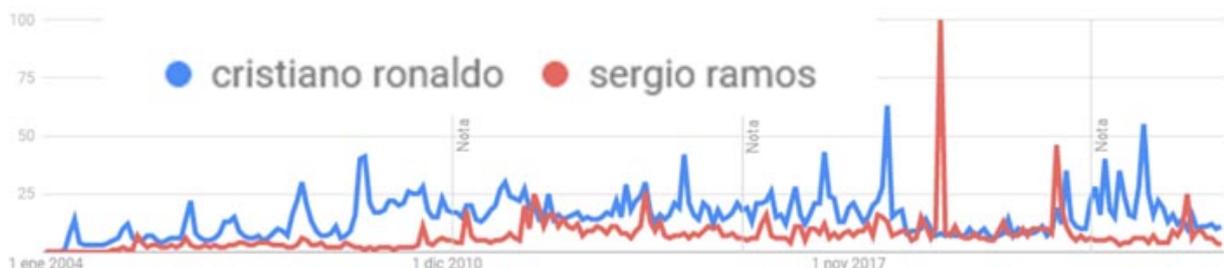
A pesar de los avances hay desafíos importantes que resolver en analítica deportiva, concretamente de acceso a los datos, aspecto que dificulta considerablemente el desarrollo de muchas investigaciones. Es por ello por lo que muchos investigadores han de trabajar con datos eventing de acceso abierto o recurrir a datos de redes sociales. Estos datos que son baratos y accesibles se han revelado como útiles por lo que representan una buena línea de investigación.

En particular, no ha pasado desapercibido el (probable) efecto de la popularidad en el valor de mercado, con el fin de predecir las tasas de transferencia. En los últimos años, los investigadores ya han incorporado medidas de popularidad como el número de seguidores en Facebook, Twitter, and Instagram, el número de links en Google cuando pones un nombre, las visitas de Wikipedia y Reddit, vídeos de YouTube y Google Trends (GT) para estimar el valor de mercado y predecir el precio del traspaso de un jugador de fútbol, demostrando que la popularidad tiene un efecto estadísticamente significativo.

Google Trends (GT)

En ese contexto, GT es una herramienta muy interesante que permite medir el interés que despierta un tema o una persona en el mundo a lo largo del tiempo. Dicha herramienta proporciona una serie temporal que mide de forma granular (por semanas, meses, años) el número de búsquedas realizadas en Google proporcionando índices normalizados dependientes del tema de 0 a 100 (figura 2).

Figura 2: Ejemplo de índices de GT.



Fuente: <https://trends.google.com/>

Además de proporcionar datos en abierto, GT presenta varias ventajas para el análisis deportivo (Malagón-Selma et al. 2023):

- Recoge una evolución temporal.
- Resume y recopila información sobre otras variables de popularidad relacionadas con

Google (por ejemplo, el número de visitas a Wikipedia, YouTube o Google).

- Ofrece un indicador que se puede utilizar como proxy de popularidad y que puede utilizarse para muchos fines (publicitarios, valoración, renovación).

Finalmente, y no por ello menos importante, cabe resaltar que el usar datos abiertos y software libre facilita la reproducibilidad y replicabilidad de los resultados, así como la confianza en las investigaciones.

En definitiva

El análisis deportivo ha evolucionado desde sus modestos comienzos en el siglo XIX hasta convertirse en una disciplina integral que combina datos, tecnología y estrategia. Con el acceso a datos abiertos y el uso de herramientas como GT, se abre un horizonte de posibilidades de datos en abierto para la investigación y el desarrollo en este fascinante campo del análisis deportivo.

Ana Debón Aucejo. Departamento de Estadística e IO Aplicadas y Calidad

Universidad Politécnica de València

andean@upv.edu.es

Referencias bibliográficas

Malagón-Selma, P., Debón, A., & Domenech, J. (2023). Measuring the popularity of football players with Google Trends. Plos one, 18(8), e0289213.

Link to Original article: <https://www.seed-deporte.es/noticias/datos-en-abierto-e-investigacion-en-futbol-del-box-score-al-google-trends?elem=318670>