Análisis de estadísticas de NBA

Máster data analytics semipresencial

Bryan Eduardo De Sousa Velásquez

Juan García Bericua Goyanes

Nuria Heras Louzao

Victor Maestro Calvo

Alberto Sánchez Delgado

2020

INDICE

# INTRODUCCIÓN

# ANTECEDENTES

# HIPOTESIS

# METODOLOGIA DE TRABAJO

# COMO HA EVOLUCIONADO EL BASKET. EVOLUCION GRAFICA

# MODELO DE EQUIPO/JUGADOR PARA SABER CUAL ES LA MEJOR VARIABLE

# MODELO PARA SABER QUIEN VA GANAR

# RESPUESTAS A LAS HIPOTESIS

# REFERENCIAS BIBLIOGRÁFICAS / WEBGRAFÍA

* Renato Amorim Torres. ***Prediction of NBA games based on Machine Learning Methods.***(2013) <http://homepages.cae.wisc.edu/~ece539/fall13/project/AmorimTorres_rpt.pdf>.
* Matthew Beckler, Hongfei Wang. **NBA Oracle** (<http://www.mbeckler.org/coursework/2008-2009/10701_report.pdf> )
* <https://towardsdatascience.com/predicting-2018-19-nbas-most-valuable-player-using-machine-learning-512e577032e3>
* <https://www.basketball-reference.com/>
* SERRANO y BAKER, Basketball Analytics, 2013, Advanced Metrics, Lake St. Louis
* <https://www.basketball-reference.com/leagues/NBA_2014.html#all_team_stats>
* <http://sportsmadeinusa.com/baloncesto/nba/ha-cambiado-el-modo-de-jugar-nba/>
* García, J. Ibáñez, S. J., Feu, S. (2010). Estadísticas de juego que discriminan las selecciones participantes en los Campeonatos del Mundo de 2006 de Baloncesto, en función del nivel y género de los equipos. Revista Kronos, IX, 17, 57-63.
* Sampaio, J., Fraga, F., Silva, T. (2004). La evolución de las estadísticas de los tiros libres en partidos de baloncesto de formación. III Congreso de la Asociación Española de Ciencias del Deporte. Valencia, 11-13 de marzo.
* Sampaio, J., Leite, N. (2006). ¿Por qué ganaron o perdieron los partidos de baloncesto los equipos que participaron en el Eurobasket 2005? Revista Kronos, nº9, pp.67-73.
* Sampaio, J., Lorenzo, A. Gómez Ruano, M. A., Matalarranha, J., Ibáñez, S. I., Ortega, E. (2009). Análisis de las estadísticas discriminantes en jugadores de baloncesto según su puesto específico, en las finales de las competiciones europeas (1988- 2006). Diferencias entre jugadores titulares y suplentes. Apunts, 96, 53-58.

*“La evaluación del jugador de baloncesto en los partidos es un proceso que despierta mucha atención de los entrenadores e investigadores” (Sampaio, 2002).*

# INTRODUCCIÓN

En un mundo en el que hasta ahora hemos basado nuestras decisiones empresariales y personales en experiencias y estadísticas de pequeños conjuntos de información, hemos pasado a un mundo en el que disponemos de gran cantidad de información.

La era actual de tecnologías de la información y comunicación, ha permitido el registro de grandes cantidades de datos para posteriormente manipular la información y sacar conclusiones sobre la conducta de los seres humanos, ventajas para empresas y con aplicaciones para todos los sectores de la sociedad.

El análisis de grandes cantidades de datos para convertirlos en información relevante resulta una tarea compleja, pero que se hay ido aplicando a todos los aspectos de la vida, siendo el deporte uno de los que más ventajas se han obtenido, especialmente de deportes de táctica y estadísticas como es el baloncesto.

La NBA está en constante evolución. El estilo de juego, las capacidades de los jugadores, la información de los equipos, son variables que han ido modificando el estilo de juego, hacia un estilo más defensivo y físico, en el que prima la planificación y la innovación, no tiene cabida los jugadores y entrenadores que son absorbidos por la gran cantidad de información que se extrae en cada partido. Dado este paradigma, nos preguntamos, ¿Cómo ha evolucionado el estilo de juego?( Intro a la parte de Nuria)

Dentro de los datos obtenidos de cada equipo, como influye cada jugador en computo de juego global del equipo. ¿Los grandes jugadores siguen teniendo la influencia en el juego que tenían antes? Cómo saben las franquciias cual es el estilo de jugador que mejor se adapta a las condiciones del equipo?

La búsqueda de una variable que cuantifique esta métrica, es uno de los objetivos que intentaremos obtener con el análisis de los datos en este trabajo ( Intro a la parte de Alberto)

El conocimiento del estilo de juego y de qué jugador es más óptimo para nuestro estilo de juego nos llega a preguntarnos si es posible la proyección de un modelos que nos permite averiguar las probabilidades y derrotas. Esta variable la hemos aplicado junto con las distancias que hay entre los campos de cada equipo y la influencia de los lesionados en esta.

# ANTECEDENTES

Todos los entrenadores e investigadores han buscado la forma de conseguir sacarle el máximo rendimiento a las estadísticas que se recogen durante el tiempo del partido, tanto de sus propios jugadores como del resto de equipos de la NBA. Dentro de los avances obtenidos en el desarrollo y aplicación de nuevos métodos para el análisis cuantitativo de datos deportivos, sobre todo utilizando técnicas estadísticas. Esa acumulación incalculable de datos requiere de conclusiones que nos permitan analizar el estado de nuestro equipo, valorar el impacto de nuestros jugadores dentro del colectivo, prever futuras líneas de actuación y hasta cambiar puntualmente el estilo de juego en función de las estadísticas propias y del equipo rival

Desde la implementación hace más de un lustro de la medición, utilización y comprensión de los datos de estadísticas convencionales como, puntos, rebotes, asistencias nos ha permitido ir conociendo cómo ha evolucionado el juego.

La incorporación de tecnología para recabar datos se ha ido realizado paulatinamente, aunque el mayor avance se produce en el año 2013, con la instalación de cámaras de registro de movimientos, para proporcionar a los entrenadores, jugadores y aficionados la información de todos los partidos.

Los principales estudios se centran en el **ritmo de juego** (se centra en el uso de las faltas, las posesiones, y las jugadas desarrolladas durante el tiempo de partido), los **porcentajes de tiro** (desarrollo de un algoritmo para conocer le valor medio de los tiros en el campo, contando con un valor extra para el lanzamiento de línea de tres) y el intento de conocer mediante el análisis de datos cual es el **mejor ataque y la mejor defensa.**

Siendo estos los temas principales, se han desarrollado varios estudios acerca de las pérdidas de balón y varios estudios sobre las votaciones para los jugadores en el draf y la evolución del juego de triples.

Entre los estudios que destacan por su notoriedad se encuentran, el estudio de Ibáñez y Col y Sampaio y Janeira ambos del 2003 en el que sus conclusiones es que los equipos que anotan más tiros de 2 puntos y recogen más balones debajo del arco obtienen un índice mayor de partidos ganados.

La tesis de Sampaio y Leite, recoge los datos del Eurobasket d 2005, en el que concluyen que las estadísticas más diferenciadora se situaba en la línea d tres, los puntos fallados y las asistencias. Con este estudio se planteaba que la situacio ofensiva se concluia con un tiro cercano, con elevadas probabilidades de aciero, para minimizar los riesgos del tiro exterior y escogiendo las mejores opciones de lanzamiento.

un estudio de Hierro (2002), tras analizar la liga ACB y la NBA, concluye que los equipos con mejores porcentajes de tiros de 2 y de 3, mayor número de faltas recibidas, tapones y asistencias, obtenían más victorias; no considerando el número de pérdidas de balón y los tiros libres como aspectos definitivos.

No todos los estudios se centran en las estadisiticas, ya que algunos toman de referencia la edad de los jugadores en la canga, (García, Ibañez y Feu (2007)), así como encontrar la posición más influyente en los partidos. Este estudio (Sampaio et Al, 2009), determinó que los jugadores más importantes eran los bases titulares y los pívots suplentes, reflejándose su aportación en los lanzamientos de 3 puntos, los rebotes y los lanzamientos de 2 puntos.

Otra de las líneas de investigación que hemos encontrado en el análisis de la literatura científica, es el estudio de las victorias y las derrotas en función de la condición del equipo como local o visitante. En la mayoría de casos, los equipos que juegan en casa, tienen mayor ventaja (Courneya y Carron, 1992; Sánchez et al., 2001; Smith, 2005), debido a factores como viajes realizados, el conocimiento de la cancha de juego, la presión percibida por los árbitros, la familiaridad con el estadio, el público y determinadas variables psicológicas

# METODOLOGÍA DE TRABAJO

Aunque hay muchos datos ricos disponibles, muchas veces no están listos para ser utilizados.

# GLOSARIO BBDD

Los primeros términos para iniciarse en las bases de datos materia deberían ser el Offensive y defensive rating, el pace o ritmo y el true shooting percentage.

* Tiros Intentados: Tiros realizados por el jugador (de campo). De campo significa todos, menos lo tiros libres.
* Puntos: Puntos anotados
* Minutos Jugados: Minutos disputados (el partido dura 48 minutos)
* Porcentaje de Tiro: Porcentaje de tiros anotados
* Rebotes: Rebotes capturados, es decir, balones capturados tras tiro fallado (de tu equipo o delcontrario)
* Tapones: Tiros del equipo contrario bloqueados
* Asistencias: Pases previos a canasta
* Recuperaciones: Balones recuperaciones (robos de balon)
* Perdidas: Balones perdidos
* Faltas: Faltas realizadas
* Tiros Fallados: Tiros fallados
* Valoracion: valoracion = (int(puntos) + int(rebotes) + int(recuperaciones) + int(tapones) + int(asistencias) ) - (int(tirosFallados) + int(Faltas) + int(Perdidas) )
* valoracionDir valoracionDir = valoracion + (int(masMenos) - int(diferencia))

# CÓMO HA EVOLUCIONADO EL JUEGO DEL BALONCESTO

<https://www.blogdebasket.com/2020/01/15/increible-asi-cambiado-zona-tiro-nba-20-anos-diferencia>

EL juego que se desarrolla actualmente en la NBA no es el mismo juego que se realizaba cuando se comenzaron a recoger datos. El análisis de datos de juego ha creado un antes y un después en el mundo del baloncesto. Tanto así que ha dado lugar a cambios importantes en la forma de jugar y las estrategias de los equipos.

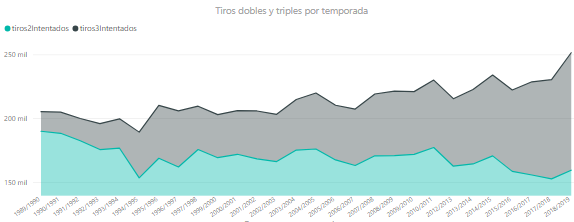
Hoy en día, cada equipo hace sus propios análisis de datos para determinar fichajes, estrategias a llevar a cabo con cada contrincante, sacar el mayor partido a cada componente del equipo, determinar minutos en cancha y partidos en los que participará cada jugador… En definitiva, toda una mina de oro para este deporte.

FORMAS EN LAS QUE SE ESTÁN OBTENIENDO DATOS

FECHA EN QUE SE COMENZÓ A ESTUDIAR LOS DATOS

La variación en la forma que se están obteniendo los datos y del año en el que comenzó a registrarse información está repercutiendo en el estilo de juego desarrollado por los equipos y en la forma de afrontar los partidos, buscando las estrategias más eficientes para conseguir el mayor número de puntos.

Con un simple vistazo a los datos de tiros por temporada a lo largo de los últimos 30 años, podemos ver que el volumen ha experimentado un gran aumento.



Esto se debe a que cada vez el juego es más rápido y por lo tanto da lugar a un mayor número de jugadas por partido. Y con ello, la oportunidad de anotar una mayor cantidad de tantos en cada encuentro.

Pero, ¿Cómo podemos conocer que el juego es más rápido? Se producen más intentos de tiro ahora que hace 20 años?

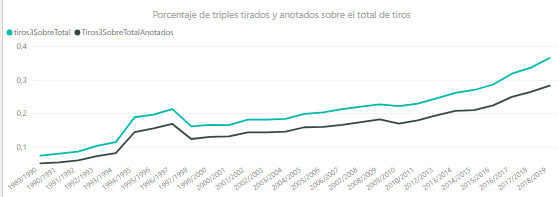
En el siguiente gráfico, comenzamos a analizar los puntos por temporada desde 1989/1990 hasta la temporada 2018/2019. La evolución se produce de forma lineal a lo largo del tiempo, teniendo un margen de crecimiento muy parecido en todos ambas metricas.



Sin embargo, el hecho de tirar más veces a canasta no es lo único que ha favorecido el aumento de puntos por temporada. Ha sido otro de los cambios tácticos importantes que ha dado la analítica de datos lo que más ha influido en este aumento.

Se trata de la importancia que se está dando a los tiros triples sobre los de 2 puntos, lo cual está invirtiendo el porcentaje de estos sobre el total de tiros, tanto en el caso de tiros intentados como en los anotados.

Como podemos ver en el siguiente gráfico, mientras en la temporada de 1989/1990, los triples sólo suponían un 7% sobre el total de tiros realizados y un 5% sobre los anotados, en la temporada de 2018/2019 estos porcentajes suponen un 37% y 28% respectivamente. Sobretodo se observa un aumento mucho más acusado en las últimas 5 temporadas, donde el incremento es de 11 y 7 puntos porcentuales respectivamente.



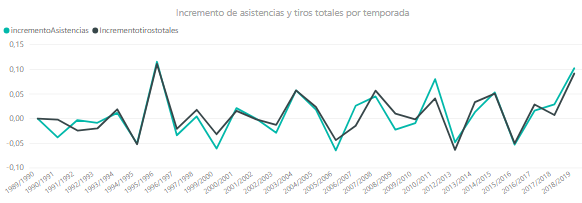
La respuesta a por qué se está llevando a cabo esta sustitución de tiros de dos puntos por tiros triples, se encuentra en la esperanza matemática que supone cada uno de ellos.

Como es evidente, el porcentaje de efectividad es superior en tiros de 2 puntos que en tiros de 3 puntos, -además este dato no ha experimentado grandes variaciones en el tiempo a pesar de que se esté tendiendo a una mayor práctica de tiros triples sobre dobles-, siendo de una media del 50% y del 35% respectivamente. Sin embargo, el hecho de anotar un triple, supone en el marcador un 50% más de puntos que un tiro doble.

Además, también se trata de una cuestión táctica. Al realizar el ataque desde fuera como es el caso de los triples, más tienen que abrirse los rivales. De esta manera resultaría más fácil atacar el aro, y por parte de la defensa, más difícil el apoyo entre los jugadores, ya que tienen que cubrir más terreno.

Una idea que podría surgir del hecho de que se esté tendiendo a un juego más basado en triples y se estén llevando a cabo jugadas más rápidas, es que el tipo de juego esté tendiendo a un modelo más individualista y menos colaborativo. Sin embargo, podemos descartar esa idea con un vistazo rápido a la evolución de los datos de tiros y asistencias en conjunto.

Como podemos observar en el siguiente gráfico, el incremento o disminución porcentual de cada temporada respecto a la anterior sigue una evolución muy similar entre ambas medidas, por lo que no podemos evidenciar un cambio en el juego hacia una tendencia más o menos individualista. Los datos de tiros totales y asistencias evolucionan en la misma dirección y en proporciones similares.



# MODELO DE EQUIPO/JUGADOR PARA SABER CUAL ES LA MEJOR VARIABLE

Con la incorporación de los datos, los equipos han logrado descubrir aquellos jugadores que por sus números no eran los mejores, pero que en el fondo desempeñaban una labor fundamental en el juego de equipo.