Universidad de Cantabria

Máster en ciencia de datos

Ciencia de datos en el deporte de élite: Seguridad, Privacidad y Ética

Autor:
Óscar Mirones Alonso

Marzo 2020

1. Introducción

Nos encontramos en una era caracterizada por la avalancha de datos. Los datos abarcan múltiples tipos y diferentes ámbitos: médicina, meteorología, economía, ciencias sociales... El conocimiento de los datos supone un potente desarrollo científico, tecnológico y económico en estos campos. Los avances de la inteligencia artificial son destacados y numerosos. De hecho, tenemos varios ejemplos con los que convivimos día a día. Simples aplicaciones como el reconocimiento facial, altavoces inteligentes (Alexa Echo Dot de Amazon) o el reconocimiento de imágenes son algunos de ejemplos. Sin embargo muchas cuestiones éticas y de seguridad se ciernen sobre el uso de los datos en cualquier campo. Un ejemplo claro en el aspecto ético es la discriminación racial por parte de algunas aplicaciones de reconocimiento facial. Otras aplicaciones pueden dar problemas de seguridad o anonimato de nuestros datos como la polémica generada por Strava I. Debido a estos hechos, la preocupación por la seguridad de los datos y las razones éticas ha ido creciendo paulatinamente en todos los terrenos de la ciencia de datos.

¿Qué sucede con el deporte?¿Ha sido afectado también por la era 'Big Data'?. El deporte no ha sido excepción y también ha sido afectado, aunque probablemente en menor medida. Destacar que en el presente trabajo solo nos referiremos al deporte de élite. El deporte a más bajo nivel no tiene los recursos económicos suficientes para tratar o almacenar adecuadamente la información de los datos. En comparación con el resto de campos, la ciencia de datos no ha tenido tanto impacto en el deporte en general. La sección más desarrollada es la Fórmula Uno, donde en cada nanosegundo se están obteniendo miles de datos. De esto hablaremos más adelante. El resto de deportes que no precisa de tanta tecnología como el fútbol, baloncesto o rugby, el desarrollo no ha sido muy destacado. Esto no quiere decir que no existan datos interesantes en estos deportes. En general, los clubes u organizaciones aún no han encontrado una forma correcta de aprovecharlos. Por otra parte, las entidades que se hacen cargo de las competiciones constantemente nos bombardean con datos. Bien sea de forma directa con estadísticas de jugadores/equipos o de forma indirecta con representaciones como los mapas de calor. Los mapas de calor son un ejemplo del uso de datos en estos deportes. Los jugadores son monitorizados y se extrae información cada minuto de juego. Además esta información puede ser variada como la velocidad máxima que se alcanza, kilómetros recorridos, zonas de preferencia ...

En las siguientes secciones analizaremos el uso de los datos y los problemas éticos o de seguridad que pueden ocurrir. Distinguiremos en los deportes tradicionales(fútbol, baloncesto o rugby) y Fórmula Uno. Esta distinción se debe a que la Fórmula Uno se ha convertido en un 'deporte de datos' y su análisis no es comparable a los demás. Nos centraremos en el problema ético en deportes tradicionales. Por otro lado, nos interesará comentar sobre la seguridad en Fórmula Uno, debido a que la ciberseguridad es un tema muy importante para el éxito en el campo y además está en constante crecimiento.

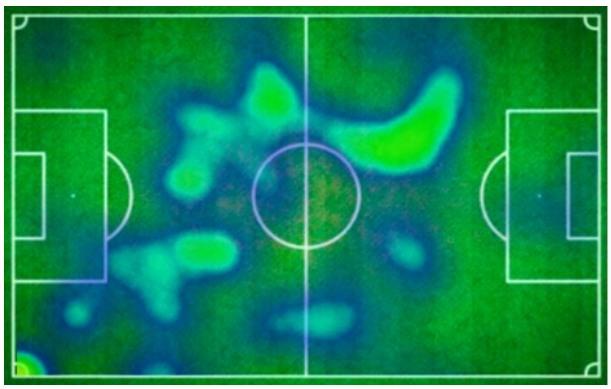

Figura 1: Mapa de calor de Casemiro durante Levante UD-Real Madrid J25. Fuente:LaLiga

2. Problemas Éticos. Seguridad

2.1. Ética en deportes tradicionales

Como mencionamos anteriormente, en la mayoría de los deportes el fenómeno de Big Data o ciencia de datos no está tan desarrollado como en otros ámbitos. Los mayoría de clubes son reticentes a utilizarlo. Sin embargo esto no quiere decir que no se esté avanzando en estos campos. En baloncesto algunos equipos están comenzando a utilizar técnicas de análisis de datos y representación para preparar sus encuentros. Un ejemplo típico es el estudio de grafos elaborados a partir de datos de los partidos del equipo rival. Se busca analizar las decisiones que hay que tomar para defender de modo que la probabilidad de que el ataque fracase sea la mayor posible.

Vamos a centrarnos en el caso concreto del fútbol para explicar los problemas éticos o de seguridad que pueden aparecer. Los clubes de élite que están utilizando datos para análisis táctico abarcan muchísima información. Desde datos históricos frente a sus oponentes hasta datos ambientales, arbitraje o factores psicológicos de los jugadores.

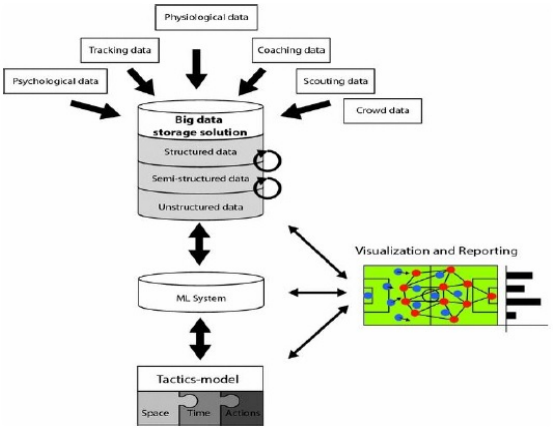

Figura 2: Tecnología Big Data para análisis táctico en fútbol [4]

Este aglutinamiento de semejante información lleva consigo un riesgo considerable. ¿Qué sucede si esta información cae en las manos equivocadas? Acceder a información privada de los jugadores como factores biológicos, psicológicos, entrenamiento específico... lleva consigo dos tipos de riesgos esencialmente:

- Desarrollar estrategias para explotar debilidades. Esto supone una mayor probabilidad de derrota, lo que implica en el alto nivel una pérdida económica y deportiva sustancial.
- Ventaja en conocimiento del mercado.

De ambos casos se pueden extraer problemas éticos y de privacidad. Comencemos analizando el primer caso. En primer lugar, el acceso a información privada sin consentimiento alguno ya es un síntoma claro de un ataque a la privacidad tanto de la institución como la de los jugadores que la componen. El conocimiento mediante estos medios de entrenamientos privados para preparar el partido es un tema que éticamente no parece correcto. Esto garantiza que existe una desigualdad a la hora de competir en duelos directos entre los equipos. Además, se atenta contra el derecho a la privacidad del individuo al conocerse factores psicológicos o biólogicos, ya que son datos sensibles. Es obvio que tradicionalmente existían ojeadores de equipos que veían partidos de sus rivales para obtener información de las tácticas. Sin embargo, este caso sería como espíar un entrenamiento a puerta cerrada donde nadie tiene acceso y se preparan estrategias específicas de cara al partido. Un caso conocido de robo de información entre equipos de MLB fue en 2015

entre los St.Louis Cardinals y los Houston Astros. Este caso estuvo bajo investigación federal, manteniéndose que los directivos de los Cardinals habían accedido ilegalmente a información propiedad de su rival de liga. Así, obtuvieron información desde scouting y análisis estadístico hasta regímenes de dieta y tratamiento psicológico. El resultado fue una penalización deportiva y 2 millones de dólares como compensación al equipo rival. Además, el directivo de scouting Chris Correa fue sancionado de por vida en la MLB y con una pena de prisión de 46 meses.

La ventaja en conocimiento de mercado lleva consigo un problema ético y de privacidad de los individuos. El problema ético reside en que con dicha información se pude generar ventaja económica en la compra y/o venta de jugadores. Esto supone una desigualdad para los demás clubes a la hora de intervenir en el mercado de traspasos. Así mismo, el uso de información privada sobre sus jugadores puede provocar desigualdades en el mercado de traspasos. Veámoslo más claro con un ejemplo: Supongamos que un club A está interesado en un jugador del club B. El club B tiene información privada sobre su jugador. Analizando parámetros psicológicos y biológicos el club dueño del jugador ha observado que ese jugador es propenso a tener una mala recuperación en las lesiones por factores extra deportivos. Esto supondría una rebaja del valor de mercado, sin embargo el club A al no tener acceso a esta información, paga el precio de mercado del jugador. Con lo cual, se ha beneficiado de esta información privada y ha obtenido ganancías gracias a ello. Mientras tanto el club A no está en igualdad de condiciones y ha sufrido una pérdida ficticia, además de fichar a un jugador que puede dar problemas. Además esto puede perjudicar gravemente al jugador. Una posible filtración de información sobre mala recuperación de lesiones, fragilidad muscular, malos hábitos o problemas psicobiológicos pueden repercutir en su trabajo. Los contratos que podría obtener serían seguramente peores en caso de que esos datos sensibles no se conociesen. Con lo cual también es perjudicial para el jugador.

Por otro lado, también puede aparecer un problema moral entre jugador y equipo. Pensemos que un equipo realiza seguimiento psicológico, físico o los hábitos diarios de un jugador monitorizándole. El equipo analiza sus datos y encuentra que ciertos jugadores pueden sufrir un alto riesgo de lesiones o una recuperación más lenta y costosa de lo habitual. Entonces analizando esa información personal el equipo decidiría sustituir a esos jugadores por otros para tener un beneficio mayor a la hora de competir. Sin embargo, los jugadores sustituidos han salido perjudicados. Además en ocasiones estas razones en el deporte de élite se suelen filtrar a los medios y tener una gran repercusión, por lo que los jugadores son aún más perjudicados todavía. Así podriamos entrar en un problema ético de utilitarianismo, es decir, la sustitución en la plantilla de esos individuos se busca para el beneficio de unos cuantos. Bien sea de todos los jugadores y cuerpo técnico para obtener un mayor rendimiento deportivo, o bien para el beneficio de la entidad deportiva en relación a los dirigentes y su accionariado. Ya que generalmente un mayor rendimiento deportivo da lugar a una mejora económica. Sin embargo, los jugadores reemplazados pueden sufrir consecuencias negativas (cómo vimos en el ejemplo del mercado de traspasos): menos ofertas en el futuro, reducción salarial en futuros contratos o incluso problemas psicológicos que puedan repercutir en un peor rendimiento (como se han demostrado algunos casos).

2.2. Fórmula Uno: Un deporte de datos. Importancia de la seguridad

Los expertos en el área afirman que la Fórmula Uno es un deporte de datos. Los coches llevan sensores que permiten examinar minuciosamente el rendimiento del vehículo cada vuelta. Las escuderías hoy en día reciben más datos de los que pueden permitirse almacenar. Como dato, Williams genera 200 Gb por carrera cada fin de semana. En esos datos, 100 Gb corresponden a los datos telemétricos del coche. El resto son datos de condiciones meteorológicas, video, neumáticos o de voz. Comparándolo con otras entidades no es una gran cantidad, pero hay que tener en cuenta que están viajando por todo el mundo y deben mantenerlos en reposo y en movimiento. Por ello, el almacenamiento es limitado.

El problema principal que existe en la Fórmula Uno en este ámbito es la seguridad de los datos. La vulnerabilidad de los datos es una preocupación en los equipos. Cada nanosegundo

de la carrera se procesan datos y se analizan para observar el rendimiento. Si se produce una caída del sistema, una pérdida de datos o incluso una infección de malware, puede repercutir en una peor actuación en la carrera. Esto lleva consigo pérdidas a nivel económico y deportivas considerables.

Los virus informáticos también suponen una amenaza considerable en este campo. En 2014, la escudería Marussia F1 sufrió un ataque de este tipo y en 4 días de ensayos solo pudieron dar 29 vueltas. Con esto se puede ver lo desastroso que puede ser un ataque de este estilo. Esto es una amenaza común. Escuderías han manifestado que pueden detectar en un mes hasta 300 instancias de malware usando filtros de email. De hecho, si algunos datos caen en manos equivocadas podría ser desastroso, razón por la cual equipos como Williams F1 se están tomando la seguridad extremadamente en serio.

Sian John, especialista en ciberseguridad, trabajó en Williams para mejorar la seguridad de los datos. Ella manifestó que se dispone de la tecnología para proteger la información en este campo: " Se trata de asegurarse de que los equipos lo pongan en marcha. Los datos se pueden proteger con contraseñas y cifrado. Se trata de poner en práctica las precauciones correctas pero haciéndolo de la manera correcta para que los datos sigan siendo utilizables". Además, Williams está llevando a cabo proyectos que priorizan la ciberseguridad sin desarrollar algoritmos de inteligencia artificial "para evitar problemas éticos", como anunció su director de operaciones Graeme Hackland.

Graeme Hackland realizó unas interesantes declaraciones donde manifestó que un gran reto de ciberseguridad para su compañía es la confianza interna y la verificación de la gente que trabaja allí. Se refiere a evitar un mal uso de información confidencial de Williams o sus clientes por parte de sus trabajadores. Es decir, se trata como una monitorización constante de sus trabajadores, incluso de forma inadvertida. Hackland razona que estas restrictivas medidas se deben a que este campo es ultracompetitivo y un mal uso de datos o de información puede llevarte al desastre en la actuación en una carrera y sus sucesivas. Aquí se podría entrar en el debate si se vulnera la privacidad de la persona. Sin embargo, desde Williams sostienen que los trabajadores que entrar en proyectos de esta índole dan su consentimiento para esta "monitorización"por parte de la entidad. Con lo cual, si se tiene el consentimiento de todos los trabajadores no hay vulnerabilidad de privacidad de la persona por ninguna parte. Para finalizar, destacar que se están llevando a cabo sesiones de educación sobre seguridad. Dichas sesiones son necesarias para que miembros de un staff que nunca han sido expuestos a estos problemas sean conscientes de la amenaza que puden llegar a ser.

Referencias

[1] Batra, S. (2017), *The Role of Data Analytics in Modern Day Sports* , Georgia Institute of Technology

[2] Grow, L. , Grow, N. (2016), *Protecting Big Data in the Big Leagues: Trade Secrets in Professional Sports*, Springer Vol.74 Article 7

[3] Marr, B. (2017), *The Big Risks Of Big Data In Sports*, Artículo disponible en:
`https://www.forbes.com/sites/bernardmarr/2017/04/28/the-big-risks-of-big-data-in-sports/#3620f3a67c6f`

[4] Memmert, D., Rein, R. (2016), *Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science*, Springer

[5] Palmer, D. (2016), *How a Formula 1 team protects itself from hackers and data breaches* Artículo disponible en: `https://www.zdnet.com/article/how-a-formula-1-team-protects-itself-from-hackers-and-data-breaches/`