

Práctica 1: Fernando de Castilla

Jugadores de todos los equipos que compiten en la LFP durante la temporada 2017/2018

El conjunto de datos alberga decenas de indicadores estadísticos relacionados con aspectos técnico-tácticos y de rendimiento de cada uno de los jugadores que forman todos los equipos que compiten en las dos máximas divisiones de la Liga de Fútbol Profesional española durante la temporada 2017/2018.



Contexto:

En un entorno cada vez más transformado digitalmente, ni siquiera el deporte queda al margen de esta disrupción tecnológica. En particular, el mundo del fútbol cada vez va abriendo más sus fronteras a la innovación tecnológica y, actualmente, existen multitud de dispositivos electrónicos encargados de registrar en video todo lo que acontece durante el transcurso del juego, monitorizar en tiempo real las condiciones meteorológicas, así como hacer un seguimiento exhaustivo y personalizado de cada uno de los jugadores (tracking). Ello ha favorecido la toma masiva de datos sobre jugadores, que posibiliten su análisis posterior, con el objeto de explotarlos y poder extraer un conocimiento clave de estos.

Contenido:

El contenido del conjunto de datos es variable, en función de la consulta que realice el usuario del script en Python. Si bien por filas disponemos de un registro por cada jugador; por columnas es posible obtener variables como las siguientes: nombre, equipo, partidos jugados, minutos jugados, partidos como titular, partidos en los que ha sido sustituido, goles marcados, goles de penalti, goles en propia puerta, goles encajados por el equipo con el jugador en el campo, tarjetas amarillas recibidas, tarjetas rojas recibidas, pases completados, pases cortos, pases largos, duelos enfrentados, duelos ganados, goles de tiro libre directo, faltas recibidas, faltas cometidas, goles dentro del área, goles fuera del área, goles de cabeza, tiros a puerta, etc.

Para dotar de cierta organización a las consultas realizadas, es posible filtrar datos de futbolistas:

- Que ocupen la misma demarcación.
- Que jueguen en primera división o en segunda división.
- Relativos a aspectos generales, disciplinarios o de eficiencia.
- Sobre una jornada en concreto, o sobre el acumulado hasta una jornada.

El periodo de tiempo de los datos siempre transcurre desde el inicio de la temporada actualmente en juego hasta la última jornada disputada; esto es, obtenemos datos desde finales de agosto de 2017 (jornada 1) hasta la jornada disputada más recientemente (el 9 de abril de 2018 fue la jornada 31 de las 38).

Los datos se recogen desde la web oficial de la LFP (Liga de Fútbol Profesional, <http://www.laliga.es>) mediante técnicas de web scraping programadas en lenguaje Python.

Agradecimientos:

Los datos que se recogen son de dominio público y pueden ser accedidos vía web. Agradecemos a la LFP que comparta parte de los datos que generan y les proveen terceras compañías de recogida y toma de datos.

Anteriormente, estas labores eran emprendidas por diarios deportivos, cuyos redactores ofrecían estadísticas de los partidos en los diarios y recopilaban un historico de varias temporadas como, por ejemplo, la guía Marca de la liga.

Inspiración:

Los datos obtenidos son útiles para los propios jugadores, y para la estrategia de negocio para entrenadores, preparadores físicos, equipos médicos, ojeadores, agentes y directivos de los clubes. Muchas cuestiones deportivas pueden ser analizadas. Por ejemplo:

- ¿Cuándo suele ser el mejor/peor momento deportivo de un jugador/equipo?
- ¿En qué fases del partido es determinante un jugador?
- ¿Cuál es la carga física que puede acumular un jugador con el objeto de prevenir lesiones?

Licencia:

La licencia seleccionada es GNU General Public License v3.0. Se trata de una licencia cuyas condiciones permiten un uso poco restringido del proyecto (uso privado y comercial, modificación, distribución, uso de patentes), a la vez que el autor no tiene responsabilidades sobre el mismo ni ofrece garantías de ningún tipo. Es una licencia adecuada para el contexto de desarrollo de un proyecto académico.

Código y dataset:

<https://github.com/fernandodecastilla/EstadisticasLFP>