## Contexto.

El conjunto de datos generado contiene la información de los partidos de cada jornada de las últimas 20 temporadas de la liga de futbol de 1ª División Española, así como datos climatológicos y de temperatura de cada encuentro.

1. Definir un título para el dataset.

Futbol y clima de la primer división temporadas 2000-2020

1. Descripción del dataset.

Cada registro del conjunto de datos generado resume en pocos atributos lo más importante de cada encuentro ligero, de cada jornada de las últimas 20 temporadas de la liga de futbol de 1ª División Española (2000-2020), así como los datos meteorológicos para la ciudad del encuentro en esa fecha. Las columnas que tiene el dataset son las siguientes:

* EquipoAnf: Nombre del equipo anfitrión.
* EquipoVis: Nombre del equipo visitante.
* ResultadoAnf: Goles del equipo anfitrión.
* ResultadoVis: Goles del equipo visitante.
* Fecha: Fecha del encuentro.
* Estadio: Nombre del estadio.
* Ciudad: Nombre de la ciudad del encuentro.
* Arbitro: Nombre del arbitro del encuentro.
* Temperatura: temperatura media del día.
* ¿Precipitaciones?

1. Representación gráfica.

Diagrama

Descripción generada automáticamente

Imagen 1: “La Liga” (Liga de primera división española de futbol), estadísticas y climatología unidas por este dataset.

1. Contenido.

Los datos futbolísticos fueron obtenidos mediante técnicas de webScraping de la página [www.bdfutbol.com](http://www.bdfutbol.com/). Después de analizar la pagina y comprobar que no dispone de Api libre (aunque si de pago https://www.bdfutbol.com/es/c/api2.html) para poder acceder a los datos, pero si ver que la pagina ofrece estos datos de manera gratuita en la web. Como observaciones, cabe mencionar que se comprobó que el robot.txt no limitaba los accesos a las rutas de donde se obtuvieron los datos y que para evitar saturar el servidor de con las peticiones, estas se hicieron usando un delay.

Ademas, los datos de los estadios (equipo, estadio, ciudad) se han recuperado de la wikipedia[[1]](#footnote-2), usando también técnicas de webscraping, creando un diccionario y añadiendo los datos usando el nombre del equipo anfitrión como nexo de unión. Estos datos también existen en bdfutbol, sin embargo se decidió, recuperar estos datos de otra fuente, con el objetivo de no sobrecargar de peticiones a la pagina.

Por otro lado, los datos referentes a la climatología se obtuvieron mediante el uso de la api de la Agencia Estatal de Meteorología (AEMet). AEMet proporciona, mediante su API, parte de la información de la cual es propietaria y que trata de aportar valor dicho servicio, tal y como se comentó en su sección de noticias:

“…

AEMET continúa en la apuesta por el desarrollo de una política de apertura de datos y transparencia dirigida a atender, dentro de sus competencias como Servicio Meteorológico Nacional, a las demandas de los ciudadanos, las entidades públicas o privadas y las empresas. En esta línea, el próximo 13 de diciembre se desarrollará una Jornada en la sede institucional de AEMET en la que se tratarán distintos aspectos de interés sobre la reutilización de datos. Para inscribirse en la jornada es necesario enviar un correo a [opendata@aemet.es](mailto:opendata@aemet.es).”

Fuente del fragmento citado: https://www.aemet.es/es/noticias/2016/11/opendata

Finalmente, para la integracion de los dos dataset nos hemos encontrado el siguiente problema, el JSON resultante tiene información recogida de cada una de las estaciones meteorológicas de España, sin embargo solo se dispone de la información discretizada por comunidad y no por ciudad (como se necesita). Sin embargo si que se tiene la información referente a la geolocalización de la estación. Esto nos ha permitido el uso de la librería de Python GeoPy para obtener la ciudad de cada estación y de esta manera poder hacer el matching de los dos ficheros.

1. Agradecimientos.

Agradecemos a [www.bdfutbol.com](http://www.bdfutbol.com/) por tener públicos los datos recogidos en esta practica.

Agradecemos a la Agencia Estatal de Meteorología que pongan a disposición pública dichos datos.

1. Inspiración.

Los datos recogidos en este dataset podrían ser utilizados tanto con fines lúdicos como lucrativos. Por ejemplo, personas aficionadas al futbol o simplemente a la estadística, podría hallar interesante este conjunto de datos con el cual, no solo consultar los datos más relevantes de un partido, sino tratar de dar respuesta mediante la minería de datos a la pregunta: ¿Podría encontrar los factores pudieron influir en las derrotas de mi equipo el año pasado?

Por otro laso, en el terreno lucrativo, se podrían elaborar modelos predictivos que ayudasen a realizar apuestas deportivas, o dentro del periodismo deportivo, la figura cada vez más extendida del analista deportivo que trata de dar claves de resultados o rachas deportivas de un equipo, entre otras muchas cosas.

1. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:
   1. Released Under CC0: Public Domain License
   2. Released Under CC BY-NC-SA 4.0 License
   3. **Released Under CC BY-SA 4.0 License**
   4. **Database released under Open Database License, individual contents under Database Contents License**
   5. Other (specified above)
   6. Unknown License
2. Código.

URL:

1. Dataset.

Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

1. Anexos:

https://www.bdfutbol.com/robots.txt

User-agent: \*

Disallow: /c/archivesko.html

Disallow: /c/archivesok.html

Disallow: /c/apiko.html

Disallow: /c/apiok.html

Disallow: /es/c/archivesko.html

Disallow: /es/c/archivesok.html

Disallow: /es/c/apiko.html

Disallow: /es/c/apiok.html

Disallow: /en/c/archivesko.html

Disallow: /en/c/archivesok.html

Disallow: /en/c/apiko.html

Disallow: /en/c/apiok.html

Disallow: /bo

1. .-https://es.wikipedia.org/wiki/Anexo:Estadios\_de\_f%C3%BAtbol\_de\_Espa%C3%B1a [↑](#footnote-ref-2)