

**Modelos estadísticos para  
la predicción de LaLiga  
2019/2020**

**TRABAJO FIN DE GRADO**



*Grado en Estadística*

ÁNGEL CARO TORRES

*Sevilla, Junio de 2020*

# Índice general

Prólogo . . . . .	III
Resumen . . . . .	V
Abstract . . . . .	VI
Índice de Figuras . . . . .	VII
Índice de Cuadros . . . . .	VIII
<b>1. Introducción</b>	<b>1</b>
1.1. La estadística en la actualidad . . . . .	2
1.1.1. La estadística aplicada al deporte . . . . .	3
1.2. Evolución de la estadística deportiva con el paso del tiempo . . . . .	4
1.2.1. Introducción al Big Data como herramienta estadística . . . . .	4
1.3. Las casas de apuestas . . . . .	5
1.3.1. Intrusión en el fútbol profesional . . . . .	5
1.3.2. Métodos de predicción y acierto . . . . .	6
<b>2. Objetivos</b>	<b>7</b>
<b>3. Creación BBDD completa</b>	<b>8</b>
3.1. Procedencia de los datos . . . . .	8
3.2. Modificaciones realizadas . . . . .	9
3.2.1. Ordenación de datos . . . . .	9
3.2.2. Cambios respecto a datos originales . . . . .	10
3.2.3. Implementación de RStudio en el tratado de datos . . . . .	10
3.3. Tabla gráfica . . . . .	11
3.3.1. Otras representaciones gráficas . . . . .	13
<b>4. Creación BBDD real</b>	<b>15</b>
4.1. Definición de éxito probabilístico . . . . .	15
4.2. Modificaciones realizadas . . . . .	16
4.3. Tabla gráfica . . . . .	16
4.3.1. Otras formas de representación gráfica . . . . .	18
<b>5. Sistema de puntuación ELO</b>	<b>19</b>
5.1. Introducción . . . . .	19
5.1.1. Un poco de historia . . . . .	19
5.1.2. Dónde podemos verlo en la actualidad . . . . .	19
5.1.3. ¿Qué deportes utilizan dicha puntuación? . . . . .	20
5.2. The “ELO” Package . . . . .	20
5.2.1. Posibilidad de ocurrir un empate en un juego . . . . .	21

5.2.2. Funciones utilizadas y qué opciones útiles nos aporta . . . . .	21
5.3. Creación BBDD puntuación ELO . . . . .	22
5.3.1. Modificaciones realizadas respecto a BBDD inicial . . . . .	23
5.3.2. Tratamiento y depuración de datos . . . . .	23
5.3.3. Funcionamiento del método . . . . .	24
5.3.4. Tablas gráficas . . . . .	24
<b>6. Modelo de predicción</b>	<b>28</b>
6.1. Introducción . . . . .	28
6.2. Objetivo principal . . . . .	28
6.3. Descripción del modelo . . . . .	29
6.3.1. Variables a tomar en cada partido . . . . .	30
6.4. Creación BBDD . . . . .	30
6.4.1. Modificaciones realizadas respecto a BBDD inicial . . . . .	31
6.4.2. Funcionamiento del modelo . . . . .	33
<b>7. Comparativa: resultados 2 a 2</b>	<b>34</b>
7.1. Real - ELO . . . . .	34
7.1.1. Tabla gráfica . . . . .	34
7.1.2. Análisis de resultados . . . . .	37
7.1.2.1. Porcentaje de acierto según temporadas . . . . .	38
7.1.2.2. Porcentaje de acierto según grupos . . . . .	39
7.2. Real - Modelo . . . . .	40
7.2.1. Análisis de resultados . . . . .	40
7.2.1.1. Porcentaje de acierto según temporadas . . . . .	41
7.2.1.2. Porcentaje de acierto según grupos . . . . .	42
7.3. ELO - Modelo . . . . .	42
7.3.1. Tabla gráfica . . . . .	42
7.3.2. Otra forma de representación gráfica . . . . .	44
<b>8. Conclusiones finales obtenidas</b>	<b>45</b>
<b>A. Apéndice: Metodologías y herramientas</b>	<b>46</b>
<b>Bibliografía</b>	<b>48</b>

# Prólogo

Este trabajo surgió gracias a un imprevisto, debido a que se planteó la idea tras la confirmación por parte de la organización de LaLiga el parón inmediato de LaLiga Santander (primera división española de fútbol) [22].

Dicho parón se llevó a cabo por el decreto de estado de alarma nacional impuesto por el Estado español [23]. Este decreto fue causado por la pandemia mundial del COVID-19, siendo afectado todo el planeta al quedar totalmente paralizado; desde las mayores ciudades a los puntos menos habitados.

Por ello, vi la ocasión de llevar a cabo un trabajo que no tiene precedente alguno a lo largo de la historia ni tampoco con ningún otro deporte u evento deportivo, puesto que la última vez que LaLiga Santander se paralizó por completo fue debido a la Guerra Civil española, pero no por una pandemia. Además, equipos como el F.C. Barcelona siguió disputando partidos de otras competiciones mientras no se jugaba la competición doméstica española [9]. Actualmente esto es impensable, puesto que el parón inmediato engloba a todo tipo de competiciones.

Según ha dictado la RFEF, LaLiga Santander se retomará, ya bien sea con o sin público, con el fin de acabar la presente temporada y no acreditar puestos europeos (Uefa Champions League y Europa League) o descensos a LaLiga SmartBank si no es por méritos propios logrados a través de la disputa de los partidos restantes que quedan por completarse.

Todo este trabajo ha sido creado a través de RStudio, gracias a la plantilla creada por Pedro Luis Luque Calvo. Lo puede localizar en [24].

*Agradecimientos:*

*Este trabajo se lo quisiera dedicar a todas aquellas personas que me han ayudado hasta el día de hoy para lograr todas las metas que me he ido proponiendo a lo largo de mi vida; profesores, amigos y por supuesto mi familia.*

*En especial se lo dedico a mis padres, a mi abuela y a mi tío Paco, que siempre han apostado por mí y me han apoyado en todo momento.*

### Información sobre el autor:

Ángel Caro Torres

- Estudiante de último curso del Grado en Estadística de la Universidad de Sevilla.
- Email: [angel14caro@gmail.com](mailto:angel14caro@gmail.com)

Para **citar este trabajo** en publicaciones, utilizar el siguiente formato:

- Caro-Torres, Á.C. 2020. Modelos estadísticos para la predicción de la liga 2019/20.[4]

Para insertar esta referencia en un fichero bibliográfico BibTeX, añadir el siguiente código:

```
@Manual{AngelCaro2020,  
title = {Modelos estadísticos para la predicción de LaLiga 2019/20},  
author = {Ángel C. Caro-Torres},  
year = "2020"  
}
```

# Resumen

Este trabajo consiste en la predicción de las últimas 11 jornadas de LaLiga Santander a partir de un modelo propio de predicción y un modelo regido por el sistema de puntuación ELO, con el fin de obtener la máxima fiabilidad posible en dichas predicciones y lograr de esa forma acertar el próximo campeón de la competición doméstica española.

El contenido del trabajo se ha dividido en 8 capítulos. En el primer capítulo se habla sobre qué efecto está teniendo la estadística en la actualidad en los distintos campos posibles. Se hace algo más de hincapié en la estadística deportiva, qué evolución ha presentado en los últimos años y lo que se espera de ella en el futuro. Además, menciona el mundo de las casas de apuestas.

El segundo capítulo engloba los objetivos generales que esperamos lograr al final del presente trabajo, y el tercer capítulo habla de la creación de nuestra base de datos (BBDD) que tomaremos como inicial.

En los capítulos 4, 5 y 6 se profundiza en los dos modelos a comparar según la fiabilidad de sus predicciones en las últimas 10 temporadas. Se explica cómo funcionan ambos modelos y sus resultados finales.

Finalmente, en los capítulos 7 y 8 se toman las conclusiones finales tras haberlos comparado previamente, despejando la incertidumbre de cuál modelo es más fiable en sus predicciones y, por tanto, es mejor herramienta estadística.

# Abstract

This document consists of the predictions of the last eleven “LaLiga Santander fixtures” from an own prediction model and a model based on ELO rating system, with the purpose to achieve the maximum success possible on these predictions and, with that, obtain the next “LaLiga Santander” champion.

The content of this document has been divided into 8 chapters. In Chapter 1, I talk about how it is affecting statistics currently in all different fields. There’s a little more emphasis in sports statistics, it’s evolution over the past years and how it might evolve over the upcoming years. Furthermore, I talk about betting shops.

Chapter 2 includes the purposes we hope to achieve at the end of this document, and Chapter 3 explains how to create the DataBase which we’ll start to work.

In Chapters 4, 5 and 6 I explain how both models work and I compare them by their hit percentage in last 10 seasons. Also, we’ll compare their results.

Finally, Chapters 7 and 8 tells us what conclusions we’ve got, clearing the doubt in what model’s obtain a better hit percentage and definitely, it’s a best statistical tool.

# Índice de figuras

3.1. Búsqueda de partidos aplazados de LaLiga Santander a través de <a href="https://www.mismarcadores.com">https://www.mismarcadores.com</a> . . . . .	10
6.1. Esquema gráfico para clasificar a cada equipo en base a nuestro modelo .	29
A.1. Creación de esquemas o gráficos de entidad - relación entre otras: disponible en <a href="https://www.draw.io/">https://www.draw.io/</a> . . . . .	46
A.2. Software estadístico RStudio <a href="https://rstudio.com">https://rstudio.com</a> . . . . .	47

# Índice de cuadros

3.1. Cabecera de la base de datos a utilizar . . . . .	12
3.2. Clasificaciones correspondientes a las 5 primeras jornadas de la temporada 1999/2000 . . . . .	13
4.1. Clasificaciones finales comprendidas entre las temporadas 2009/2010 y 2018/2019 . . . . .	17
5.1. Asignación valor Gana.A en base a la probabilidad de ganar un encuentro	24
5.2. Cabecera base de datos sistema de puntuación ELO . . . . .	25
5.3. Clasificaciones finales obtenidas a través del sistema de puntuación ELO	26
6.1. Cabecera base de datos modelo de predicción . . . . .	32
6.2. Matriz de confusión obtenida . . . . .	33
7.1. Clasificaciones reales - puntuación ELO desde la temporada 2009/10 hasta la 2013/14 . . . . .	35
7.2. Clasificaciones reales - puntuación ELO desde la temporada 2014/15 hasta la 2018/19 . . . . .	36
7.3. Porcentaje de acierto del sistema de puntuación ELO . . . . .	37
7.4. Porcentaje de acierto de nuestro modelo de predicción . . . . .	40
7.5. Clasificaciones obtenidas temporada 2019/2020 . . . . .	43

# Capítulo 1

## Introducción

Conocemos a la estadística como una de las más importantes ramas de las matemáticas, puesto que poco a poco se está abriendo paso en nuestra sociedad y cada vez con más fuerza. Nombramos a la estadística como la ciencia que se encarga de la recopilación y almacenamiento de una gran masa de datos para su posterior estudio y sacar las conclusiones que sean oportunas. No tiene límites, es decir, no tiene ningún campo de estudio al que se le resista, puesto que cualquier ámbito puede transformarse a una forma numérica, bien sea como variables discretas, continuas o mixtas. Puede encontrarse en géneros opuestos, desde un carácter demográfico para el estudio de la población hasta un carácter sanitario, donde podemos aplicar tests de hipótesis para contrastar la probabilidad de que un paciente presente o no cáncer. No olvidar algunas aplicaciones en nuestro día a día como es para uso académico, en empresas y organizaciones, eventos deportivos, medios de comunicación o dispositivos electrónicos.

¿Qué significa la palabra *estadística*? ¿cuál es su origen? Como bien define el INE [19], la palabra **Estadística** proviene del alemán *Statistik*, puesto que como primer y único objetivo se tenía el estudio del Estado, con el fin de recopilar información y conocer a su población, para tomar decisiones como la cuota de impuestos, el reparto de tierras y bienes, o bien como medida de reclutamiento de soldados. De ahí surgieron las primeras formas de conocer a la población del Estado.

Como primera opción o herramienta estadística por la que se optó para conocer la población de un Estado fue el censo demográfico, la cual se planteó para obtener información de cada uno de los individuos, pero se dieron cuenta rápidamente de que era un recurso altamente costoso y laborioso. Sabiendo esto, se procedió al estudio de una pequeña parte de la población, conocido como muestra, y posteriormente se generalizaban los resultados obtenidos a todo el resto de la población.

Otras maneras de agilizar el recuento de datos fue el incorporar listas donde inscribirse nuevos fenómenos demográficos, como son los nacimientos, las defunciones o los matrimonios. De esa manera la población se mantenía actualizada sin requerir más recursos, como es el personal necesario para ir a cada domicilio pidiendo los nuevos registros en caso de haberlos.

Más adelante en el tiempo se fue añadiendo más y más preguntas que el Estado se hacía sobre su población, por lo que dió comienzo a la realización de las primeras encuestas. Éstas se basaban en distintos temas sociales, como preguntar por las prestaciones recibidas en el municipio o ciudad como son la limpieza de calles, mantención de alumbrado o si el municipio está bien dotado de supermercados/tiendas para los habitantes del mismo. Actualmente se ha ido desarrollando de una forma mucho más profundizada, dando lugar a muchas más organizaciones u empresas en la que nos encontramos registrados con nuestros datos personales.

Como curiosidad, podemos tomar como referencia para medir la importancia que ha obtenido la estadística en la historia por la presencia de ésta en algunas civilizaciones antiguas, como son Roma, China o Egipto, donde queda reflejada en aspectos religiosos. Este es el caso de la Biblia, donde se aplicaba un recuento de los varones de mayor de 20 años en el libro de los Números y así hacer constancia del número de varones que se encontraban en disposición de poder ir a la guerra.

## 1.1. La estadística en la actualidad

La estadística la tenemos muy presente durante cada día desde hace muchos años. Esto se cumple con casi toda probabilidad puesto que toda la población está registrada en las listas que hemos nombrado anteriormente; bien en el registro de nacimientos, matrimonios, en el padrón de la ciudad donde vivimos, los datos que aportamos a una empresa al darnos de alta como trabajadores, o los datos para hacernos una cuenta corriente bancaria.

Todo esto sin hablar de lo más evidente y que más datos genera de nosotros mismos; nuestros dispositivos móviles. Los teléfonos cada día son más innovadores y con más funciones, lo que equivale a generar y recibir más datos sobre nosotros cada momento en el que están en funcionamiento.

Si incluimos el ámbito formativo y empresarial, podemos asegurar el gran auge que está teniendo y seguirá teniendo posteriormente. Las empresas presentan cada vez más demanda en el manejo de grandes cantidades de datos para poder tomar todo tipo de decisiones, logrando así una mayor eficacia y efectividad en la empresa, desde la más simple hasta la más compleja decisión. Por ello, están buscándose un hueco en esta pequeña burbuja laboral cada vez más personas, fomentando así un mayor estudio y profundización en el conocimiento del Data Science, Deep Learning, Artificial Intelligence y en el Machine Learning.

Esta es una de las principales bazas que presenta la estadística, ser polivalente en cualquier estudio y grupo laboral. Al ocurrir esto, las propias empresas son conscientes de lo útiles que pueden llegar a ser los analistas o Data Science al sacar el máximo beneficio de sus organizaciones siempre tomando el mínimo riesgo posible. Por ello la estadística es una de las ramas con menos porcentaje de desempleo en nuestro país y por todo el mundo [10].

### 1.1.1. La estadística aplicada al deporte

La estadística deportiva aumenta cada año al parecer sin límite alguno. Ha pasado de ser un hobby o pasatiempo de la sociedad a ser uno de las principales fuentes de ocio y entretenimiento, así como un factor económico bastante importante en la economía de muchos países.

Actualmente todo deporte queda reflejado en sus estadísticas; desde los dardos o la petanca hasta el hockey americano, la NBA, o el fútbol europeo. Todo jugador o aficionado le gusta ver las estadísticas de su equipo, ya bien sean clasificaciones, torneos, puntos o goles anotados, rachas. . . un sinfín de variables y cuestiones que le surgen a los seguidores y cada vez está moviendo los intereses de más personas.

Por ello, la estadística deportiva no afecta únicamente a los deportistas en sí, sino también a otro subgrupo de la población muy interesada; los medios de comunicación, debido a que la retransmisión de ciertos eventos deportivos generan masas de audiencia muy por encima de la media diaria y, con ello, unos niveles de beneficios muy superiores al que suelen conseguir, además de añadir un otro gran margen de beneficios con las empresas y organizaciones que quieren publicitarse ante tales cantidades de telespectadores.

No olvidemos la cantidad de dinero que mueve el deporte, estamos hablando de cifras multimillonarias en todo el mundo, puesto que los productos deportivos de los equipos suelen ser muy solicitados por parte de los aficionados, así como las entradas de los partidos o eventos donde participen sus equipos, llegando a influir considerablemente en la economía en el PIB nacional de algunos países como son España o Alemania, datos contrastados en [6].

Hablando de futuro podríamos asegurar que la estadística irá cada vez más de la mano del mundo del deporte, así como cualquier evento deportivo de pequeña o gran escala. Realmente no es un objetivo que se marque la estadística, sino que, como anteriormente se explica, los propios aficionados y medios solicitarán cada vez más y con más matices en el juego.

Finalmente, indicar en que esta situación de la pandemia mundial del COVID-19 ha provocado la suspensión o aplazamiento de la totalidad de eventos deportivos, bien sean o no profesionales y sin importar qué tipo de deporte sea. Algunos de los casos más importantes han sido el aplazamiento del fútbol europeo, el cual no se detenía indefinidamente como es el caso de España desde la Guerra Civil española, o la suspensión sin fecha de vuelta de la NBA, otro de los mayores eventos deportivos del mundo; así como el Roland Garros de París, uno de los torneos más prestigiosos del tenis a nivel global, la EURO 2020 que se iba a celebrar el próximo verano o los Juegos Olímpicos (JJOO) que iban a celebrarse en Tokio en los meses veraniegos de julio y agosto.

## 1.2. Evolución de la estadística deportiva con el paso del tiempo

La recopilación de datos en los partidos o juegos de cierto deporte se comenzaron a tratar con los primeros eventos deportivos, con el fin de conseguir establecer lo que se conoció como clasificación u orden que tomaban los distintos equipos en función a los resultados que iban obteniendo a lo largo de dicho evento. Con situaciones como la igualdad de puntuación entre dos o más equipos ya se inició a una recolección de datos más exhaustiva, tomando más criterios como modo de desempate. De esta forma se fue ajustando cada vez más las condiciones de desempate hasta llegar a las clasificaciones actuales que están establecidas en cada deporte, cada una con sus propios criterios de orden de equipos o jugadores.

Por otro lado, actualmente los jugadores o atletas se preocupan mucho de sus propias estadísticas, puesto que ya están al alcance de los seguidores y sirve como herramienta comparativa entre los jugadores de un equipo y otro, siendo una forma de expresar las diferencias físicas entre ellos.

A día de hoy, como se ha explicado anteriormente, esta faceta de la estadística a gran escala en el deporte está en auge, por lo que cada vez hay más competencia y número de personas formadas o se están formando en este aspecto para dedicarse íntegramente en el deporte como trabajo propio. Una de las herramientas más utilizadas y con más futuro es la conocida como Big Data, siendo más que útil en muchas situaciones deportivas.

### 1.2.1. Introducción al Big Data como herramienta estadística

¿Merece la pena invertir en dicha técnica para un proyecto deportivo importante? La respuesta es sí, puesto que el/los país(es) organizador(es) de un evento a gran escala como son los JJOO o la EURO 2020 no sólo deben de supervisar los diferentes deportes o equipos, sino también los cientos de medios de comunicación o los cientos de miles de aficionados que albergan dichos eventos; además de tener que hacer previsiones de cuánto nivel de turismo se espera, niveles de ingresos a conseguir para saber hasta dónde pueden llegar en los gastos de las diferentes infraestructuras... y muchísimas más variables a estudiar.

Ahora bien, ¿realmente se han obtenido éxitos gracias a esta herramienta estadística? Como cualquier ciencia, no siempre es exacta y no siempre funciona bien, como era de esperar. Eso sí, si se aplica correctamente y se extrae la mayor información posible de los datos obtenidos, podemos asegurar que se tomará el mínimo de riesgo posible en todas las decisiones a tomar, desde qué jugadores son más propensos a lesionarse hasta la estructuración de un club u organización.

En el caso de LaLiga Santander, una forma de tomar todos los datos acerca de cada uno de los jugadores es mediante los conocidos como chalecos GPS [29], los cuales son utilizados por los jugadores durante los entrenamientos y partidos que disputan, y nos ofrecen datos sobre su rendimiento, pulsaciones o la distancia recorrida en el período de la realización física. Estos chalecos generan miles de datos por cada ejercicio físico, aportando una gran oportunidad a aquellos trabajadores dedicados al Big Data deportivo.

## 1.3. Las casas de apuestas

Las casas de apuestas han existido desde el principio de los tiempos, puesto que se originaron en la época del imperio romano [5]. En aquel entonces se realizaban apuestas entre las luchas de gladiadores, o bien en las carreras de cuádrigas, que se celebraban cada 4 años. Sin embargo, la gran mayoría de gente da su origen en la Inglaterra del siglo XVIII con las apuestas en carreras de caballos, puesto que presentaba una metodología de apuestas moderna, siendo ésta mucho más parecida a lo que entendemos actualmente como apuestas deportivas. Estas carreras fueron de lo más cotizado junto a las carreras de galgos, otro gran atractivo en ese entonces, lo que comenzaba a ser el auténtico negocio que conocemos a día de hoy.

Durante todo este tiempo hemos estado ignorando una gran fuente de ingresos en los eventos deportivos como son las casas de apuestas. Estas empresas se encargan de crear una serie de cuotas en una lista determinada de posibles apuestas a realizar, bien sea desde qué equipo ganará el próximo partido hasta, por ejemplo en el caso del fútbol, cuántas faltas se realizarán en el primer tiempo o cuántos saques de esquina habrá para el equipo visitante.

Gran parte de los seguidores o aficionados al deporte realizan apuestas deportivas. Un gran “punto a favor” de las casas de apuestas es que tienen cuotas para todo tipo de deporte para intentar lograr el máximo de personas apostantes y así aumentar sus beneficios. Esto puede darse porque realmente se puede apostar por casi todo lo que puedas pensar, siendo muy criticado en algunos casos como sucedió en el pasado mes de marzo.

Una vez comenzada la cuarentena nacional por el COVID-19, unas de las sugerencias que se llevaron a cabo con el fin del entretenimiento fue un torneo benéfico de videojuegos online organizado por el comentarista profesional Ibai Llanos [28]. El torneo fue retransmitido vía online por todo el mundo, además de numerosos canales de televisión como es Gol Tv o Movistar + LaLiga. El objetivo de éste era la recolección de donaciones voluntarias de los telespectadores, yendo todo este dinero destinado íntegramente a la lucha por el COVID-19. Al ser un torneo que atraía a bastante masa social, se aprovecharon de ello añadiendo cuotas sobre dicho torneo en sus respectivas páginas web [2].

Gracias a las tecnologías, cada vez más automatizadas en nuestras vidas, cualquier persona mayor de 18 años que tenga un dispositivo con conexión a internet o bien una sede cercana de las distintas casas que existen puede realizar apuestas deportivas. Si hablamos de organización y beneficios está muy bien llevado a cabo, aunque han recibido muchas críticas por parte de la sociedad, puesto que promueve continuamente la ludopatía cada vez con más fuerza hacia los hogares.

### 1.3.1. Intrusión en el fútbol profesional

Como no podía ser de otra forma, las casas de apuestas ya han conseguido un puesto privilegiado dentro de la élite mundial. La gran mayoría de la población, sea seguidor o no de algún deporte, sabría nombrar al menos a algunas de estas franquicias. A día de hoy, las casas de apuestas más importantes y con mayor impacto socio-económico son Bet365, William Hill o 888Sport entre otras [21]. Como es de esperar, hay muchísimas opciones para realizar apuestas, habiendo mucha competencia de mercado y, con ello, muchas más cuotas donde elegir.

El fútbol profesional no iba a ser menos. Se ha creado en los últimos años una red de mercado impactante, puesto que han logrado colarse en las pantallas de toda la población mediante publicidad y patrocinios con los distintos medios o equipos. Este ha sido el caso de los equipos profesionales de fútbol que militan en LaLiga Santander y LaLiga SmartBank, donde actualmente las distintas casas de apuestas tienen contratos millonarios con el 95 % (19/20) de los equipos de LaLiga Santander y con el 82 % (18/22) de los equipos de LaLiga SmartBank; es decir, un promedio del 88'5 % de los equipos profesionales del fútbol español, siendo ésta una de las más potentes y con más seguidores del mundo [7]. Dichos patrocinios suponen un alto nivel de ingresos en la mayoría de clubes, donde el pico más alto se lo llevan el Real Madrid C.F. y el F.C. Barcelona, ingresando entre 6 y 8 millones de euros por temporada [26]. El único equipo de 1ª división sin ningún acuerdo con las casas de apuestas es la Real Sociedad de Fútbol, que tras hacerse una encuesta entre sus abonados sobre si querían o no ser participantes de este negocio, mostró una negación rotunda de esta opción con una abrumadora mayoría del 86 % de los votos [27]. La fuente de ingresos de los clubes son muchas más y, por lo tanto, no son indispensables; eso sí, nunca está de más tener otra entrada de beneficios para mejorar las infraestructuras del club o poder realizar mejores fichajes para intentar competir lo mejor posible contra los mejores equipos del país.

Por otra parte, quedan los medios como canales de televisión o prensa, donde también existen acuerdos con las casas de apuestas. Se da en la mayoría de los partidos de cada jornada, cuando aparecen en pantalla las cuotas disponibles para ese partido. Esto ocurre en todas las grandes competiciones del mundo, como es el caso de la UEFA Champions League (UCL), competición por excelencia del viejo continente donde se enfrentan los mejores clubes de Europa y genera aún mucho más expectación de lo que causan las diferentes ligas domésticas. Tras el parón indefinido causado por el COVID-19, este sector ha quedado bastante tocado tras una serie de años de crecimiento imparable. Únicamente en el año 2018 se registraron un total de 355 millones de euros gastados en apuestas deportivas, lo que ha conllevado este año a una caída del 2 % del PIB nacional, con unas pérdidas estimadas de unos 60 millones de euros. Además, debemos añadir los más de 47000 puestos de trabajo que se encuentran en el aire al no saber cuándo se retomarán las competiciones [11].

### 1.3.2. Métodos de predicción y acierto

Ahora bien, ¿son tan fiables las casas de apuestas como para tener asegurado obtener beneficios en la mayoría de los casos? Es evidente que detrás de las casas de apuestas existen una serie de analistas de datos que ofrecen las distintas cuotas, además de poseer muchísimos más datos de los que puede obtener cualquier persona desde un dispositivo con conexión a internet. Estos analistas toman una serie de modelos de predicción que les proporciona esa fortaleza a las casas de apuestas para realizar cuotas sin mucho margen de riesgo en pérdidas de beneficios. Según un trabajo de fin de grado de la Universidad de Carlos III de Madrid, las casas de apuestas presentan un índice de acierto del 53 %, dejando claro que no es una vía fiable para obtener beneficios para uno mismo [3]. No obstante, se han dado anteriormente casos donde algunos seguidores han sido más inteligentes que las casas de apuestas y le han ganado la partida en esa ocasión. Esto sucedió en Japón, donde el investigador de la Universidad de Tokio Lisandro Kaunitz [13] logró encontrar un método de obtener beneficios sin precedente alguno. Una vez que las casas de apuestas observaron su índice de acierto, le prohibieron seguir apostando.

# Capítulo 2

## Objetivos

A lo largo de este trabajo tendremos varios pasos u objetivos a cumplir, que iremos logrando según vayan transcurriendo los distintos capítulos. La mayoría de ellos se basan en conseguir crear bases de datos útiles para la predicción de los partidos que quedan por disputarse.

Como objetivo principal tenemos conseguir que uno de nuestros modelos, bien sea el regido por el sistema de puntuación ELO o bien el modelo que crearemos nosotros mismos, estime con cierta fiabilidad el próximo campeón de LaLiga Santander, así como los equipos que le acompañen en la Uefa Champions League, Europa League y descensos; todo esto en base a los datos de las clasificaciones de las 10 temporadas anteriores (2009/2010). Dentro de este objetivo nos podemos centrar en varios subobjetivos:

- Obtener los datos que se remontan a la temporada 2009/2010 para la creación de nuestra primera base de datos, la que utilizaremos para establecer las clasificaciones estimadas por parte del paquete ELO de RStudio.
- Obtener los datos que se remontan a la temporada 1999/2000 para la creación de la base de datos que nos servirá para obtener las clasificaciones estimadas desde la temporada 2009/2010 hasta la actual y para a través de ellas obtener la base de datos que utilizaremos para el modelo que crearemos (notar que esta base de datos posee una mucha mayor escala que la base de datos utilizada para el ELO).
- Comparar ambos modelos en función a las 10 clasificaciones que obtendremos de ambos, con el fin de conseguir que nuestro modelo muestre un índice o probabilidad media de acierto lo más próximo posible al logrado por el ELO, lo cual se representará a través de distintos gráficos variados.
- Conocer más profundamente el paquete ELO de RStudio, conocer qué variantes y alternativas nos ofrece, y cuán fiable es para nuestro trabajo.
- Darnos cuenta de la imposibilidad que existe de poder predecir la mayoría de los partidos que se disputen, puesto que después de todo, tanto en el fútbol como cualquier deporte, son dos equipos en igualdad de condiciones. También existen los conocidos como casos aislados, donde el equipo a priori débil es el que sale vencedor de un partido donde presentaba opciones ínfimas de ganar.

# Capítulo 3

## Creación BBDD completa

Para la realización de este trabajo se ha requerido una base de datos bastante grande, puesto que para poder hacer unas estimaciones medianamente fiables en los partidos que quedan por disputarse de LaLiga Santander es necesario tener bastantes precedentes de los mismos y así suponer el comportamiento futuro que mantendrán dichos partidos.

Como era de esperar, no existe tal base de datos que abastezca a todos los criterios que se tendrán en cuenta en los modelos que se utilizarán en los siguientes capítulos, como son el sistema de puntuación ELO y nuestro propio modelo de predicción que crearemos más adelante. De esta forma no hemos tenido otra alternativa que la de crear nuestra propia base de datos, con la que obtendremos la información de todas las variables requeridas por ambos modelos de predicción.

### 3.1. Procedencia de los datos

El primer obstáculo para la realización de nuestra base de datos con el que nos hemos topado ha sido no poder encontrar los datos de las distintas variables de manera separada (puesto que conjunta no hay ni una sólo base de datos a la que podamos tomar como referencia en una única fuente de datos), es decir, en distintas páginas web.

La fuente de datos por la que hemos optado para tomar sus datos ha sido la recogida en [17], donde hay datos de todas las grandes ligas europeas desde hace más de 25 años en algunos casos, así como registros de resultados de las selecciones nacionales en los distintos campeonatos en los que han sido partícipes durante todos estos años atrás.

En nuestro caso se han descargado los archivos correspondientes a todos los encuentros disputados de la liga doméstica española desde la temporada 1999/2000 hasta la temporada actual, ambas inclusive (20 temporadas completas y lo disputado de la actual). Los archivos tienen originalmente un formato `.csv`, es decir, “*comma separated values*”, lo cual es perfecto para su lectura en RStudio.

Esto nos genera una base de datos global de 7870 encuentros disputados en la primera división española, además de albergar 42 equipos distintos a lo largo de estas 21 temporadas, incluyendo a los 5 equipos que han logrado proclamarse como campeón de la competición (Real Madrid, Barcelona, Atlético de Madrid, Valencia y Deportivo de la Coruña). De cada partido registrado se han obtenido hasta 105 variables distintas, de las cuales hemos seleccionado las que nos resultarán útiles posteriormente.

## 3.2. Modificaciones realizadas

Una vez obtenidos los datos, todas las operaciones a realizar se llevarán a cabo con el software estadístico RStudio. Tras la lectura de los archivos, únicamente vamos a quedarnos con las columnas correspondientes a las variables:

- **Jornada:** se denotará como variable *Jornada* a la columna donde nos proporciona la etapa de las 38 posibles que se disputan a lo largo de cada temporada.
- **Equipo local:** esta columna se corresponde a los equipos que están disputando el encuentro ante su afición en su estadio, lo que denominaremos como “*jugar en casa*”.
- **Equipo visitante:** denotaremos como *equipo visitante* al conjunto que dispute un encuentro en el estadio de su adversario. A este fenómeno lo reconoceremos como “*jugar fuera de casa*”.
- **Gol a favor local:** llamamos *gol a favor local* a la columna correspondiente en la que se insertan los goles anotados por parte del equipo local del encuentro.
- **Gol a favor visitante:** conocemos como *gol a favor visitante* al fenómeno que ocurre cuando el equipo visitante del encuentro logra anotar un gol.

Una vez obtenidos esa pequeña parte de la base de datos, nos centramos en una serie de modificaciones necesarias para poder obtener valiosa información de la misma.

### 3.2.1. Ordenación de datos

El siguiente problema con el que nos hemos encontrado ha sido el orden con el que los datos se veían reflejados tras descargarlos. Por defecto, la ordenación de los encuentros estaba establecido de forma cronológica, es decir, teniendo en cuenta únicamente la fecha de la disputa del partido.

A priori parece que es bueno, que nos puede llegar a ser útil, pero en nuestro caso no viene bien puesto que necesitaremos la clasificación de los equipos en cada una de las jornadas durante las 21 temporadas que tendremos en nuestra base de datos. Por ello, hemos recurrido a la búsqueda de aquellos partidos que fueron aplazados y, por tanto, no se jugaron para cuando se tenía programado.

Para ello hemos utilizado la aplicación gratuita de resultados deportivos *MisMarcadores*, la cual abarca desde deportes como el fútbol hasta waterpolo. Como podemos ver en la siguiente imagen (Figura 3.1), seleccionamos la temporada requerida (en nuestro caso hemos tomado la actual como ejemplo), de ahí nos vamos a la pestaña de *Resultados* donde podemos observar qué partidos se encuentran fuera de lugar entre jornadas posteriores a la misma.

Nótese que este paso ha sido algo complejo y costoso, puesto que lo ideal sería que no hubiesen partidos aplazados durante las temporadas, aunque por desgracia rara vez ocurre esto y en la gran mayoría de ocasiones nos encontramos con varios encuentros aplazados, o en algunos casos incluso jornadas completas que se disputaron en fechas que no le correspondían. De esta forma, con esta herramienta que hemos utilizado se han logrado obtener los encuentros jornada a jornada para su posterior uso en nuestro modelo de predicción.

### Búsqueda de partidos aplazados

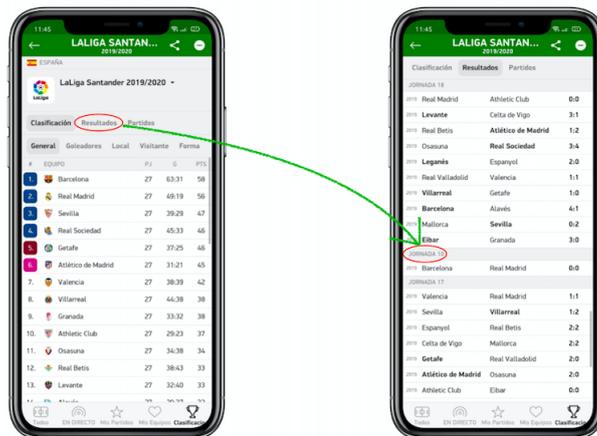


Figura 3.1: Búsqueda de partidos aplazados de LaLiga Santander a través de <https://www.mismarcadores.com>

### 3.2.2. Cambios respecto a datos originales

Una vez tenemos los archivos descargados e importados a RStudio, nos hemos quedado con las columnas requeridas y para finalizar hemos ordenado los datos correctamente. Ahora viene uno de los pasos más importantes de nuestra creación de la base de datos; la obtención de las respectivas columnas necesarias para poder aplicarlas en cada una de las variables de nuestros modelos de predicción.

### 3.2.3. Implementación de RStudio en el tratado de datos

Las columnas o variables que crearemos son varias, siempre teniendo la precaución de varios factores, como son no alterar los datos originales de las columnas principales a través de las modificaciones realizadas en las variables nuevas, o también con la opción de poder encontrarnos con el caso de añadir columnas innecesarias o redundantes en nuestros modelos, siendo en muchos casos contraproducente para nuestro estudio y favoreciendo a posibles errores cuando se proceda a la realización de consultas en nuestra base de datos.

Además de las columnas conocidas anteriormente, como son *Jornada*, *Equipo Local*, *Equipo Visitante*, *Gol a favor local* y *Gol a favor visitante*, se han añadido:

- **Temporada:** la columna *Temporada* nos facilitará en qué período se disputó cada uno de los encuentros, puesto que dos mismos equipos se han podido enfrentar muchas veces entre ellos pero se encuentran en temporadas diferentes.
- **Mitad:** al igual que *Temporada*, esta columna nos ayuda a identificar cuándo se disputó un encuentro en concreto, pero en este caso se diferencia entre *Ida* o *Vuelta*, en función de si la jornada se encuentre entre las jornadas 1 y 19 inclusives, o es perteneciente a las jornadas 20 a la 38.
- **Fecha:** la columna *Fecha* nos proporcionará la información sobre qué día de la semana y en qué mes se disputaron cada uno de los encuentros, pudiendo ser información bastante útil más adelante para los modelos de predicción.

- **Puntuación:** esta variable se compone de dos columnas diferentes, pero con el mismo objetivo. Aquí se presentan las puntuaciones acumuladas que van obteniendo los distintos equipos conforme van disputando encuentros. Se dividirá en *Puntuación Local* y *Puntuación Visitante*.
- **Goles a favor:** a diferencia de las columnas anteriores donde se anotaban los goles por parte de ambos combinados en un partido en concreto, en estas columnas se irán añadiendo los *Goles a favor* de todos los equipos según vayan avanzando en la competición. Se dividirá en *Goles a favor acumulados local* y *Goles a favor acumulados visitante*.
- **Goles en contra:** las columnas referentes a *Goles en contra* difieren de las columnas de *Goles a favor* en que éstas son completadas a través de los goles que los conjuntos van encajando con el paso de los partidos disputados. Se dividirá en *Goles en contra acumulados local* y *Goles en contra acumulados visitante*.
- **Diferencia de goles:** para este caso se completarán ambas columnas con el resultado de la diferencia entre *Goles a favor* y *Goles en contra* para cada uno de los equipos, siendo posible obtener valores negativos si un equipo en concreto ha encajado más goles que los anotados por su parte. Se dividirá en *Diferencia de goles acumulados local* y *Diferencia de goles acumulados visitante*.
- **Ranking:** por último añadiremos las columnas referentes a la posición que toman los distintos equipos en cada una de las jornadas de todas las temporadas estudiadas. Estas posiciones son las que determinarán si un equipo se encuentra en zona de descenso, media tabla, zona europea... etc. Se dividirá en *Ranking local* y *Ranking visitante*.

### 3.3. Tabla gráfica

Tras añadir todas estas nuevas variables obtenemos una base de datos de 18 columnas, 7870 filas o registros, y un total de 141.660 celdas registradas y listas para analizar. Aquí mostramos la cabecera y primeras filas de la misma:

En este caso nos hemos decidido por mostrar las primeras 5 jornadas de la temporada 1999/2000, la primera de las 21 temporadas que complementan nuestra base de datos. Hemos seleccionado este tramo puesto que son los primeros registros que tenemos almacenados cronológicamente, pudiendo sacar información a simple vista como que el Rayo Vallecano, tras las primeras 4 jornadas disputadas obtenía 12 puntos de 12 posibles, siendo líder de la clasificación; otro caso sería el Valencia, también tras la disputa de las 4 primeras jornadas, ha cosechado 4 derrotas consecutivas obteniendo un total de 0 puntos y siendo colista (última posición) en la clasificación.

Cuadro 3.1: Cabecera de la base de datos a utilizar

Temporada	Mitad	Jornada	Fecha	Local	Visitante	GoL_Loc	GoL_Vis	Punt_Loc	Punt_Vis	GF_Loc	GF_Vis	GC_Loc	GC_Vis	DG_Loc	DG_Vis	rank_Loc	rank_Vis
99_00	Ida	1	21/08/99	Ath Bilbao	Betis	1	0	3	0	1	0	0	1	1	8	15	14
99_00	Ida	1	21/08/99	Mallorca	Real Madrid	1	2	0	3	1	2	2	1	-1	1	12	4
99_00	Ida	1	21/08/99	Valencia	Santander	1	2	0	3	1	2	2	1	-1	1	13	5
99_00	Ida	1	22/08/99	Ath Madrid	Vallecano	0	2	0	3	0	2	2	0	-2	19	2	3
99_00	Ida	1	22/08/99	Barcelona	Zaragoza	2	0	3	0	2	0	0	2	2	3	18	19
99_00	Ida	1	22/08/99	La Coruna	Alaves	4	1	3	0	4	1	1	4	3	-3	1	20
99_00	Ida	1	22/08/99	Malaga	Espanol	1	0	3	0	1	0	0	1	1	6	7	16
99_00	Ida	1	22/08/99	Numancia	Valladolid	1	0	3	0	1	0	0	1	1	7	14	17
99_00	Ida	1	22/08/99	Oviedo	Celta	1	0	3	0	1	0	0	1	1	-1	17	15
99_00	Ida	1	22/08/99	Sevilla	Sociedad	2	2	1	1	2	2	2	2	0	11	10	11
99_00	Ida	2	28/08/99	Sociedad	Ath Madrid	4	1	4	0	6	1	3	6	3	-5	4	20
99_00	Ida	2	29/08/99	Alaves	Malaga	2	1	3	3	2	5	2	5	-2	0	13	10
99_00	Ida	2	29/08/99	Betis	La Coruna	0	0	1	4	0	4	1	1	-1	3	17	5
99_00	Ida	2	29/08/99	Celta	Sevilla	2	1	3	1	2	3	2	4	0	-1	9	16
99_00	Ida	2	29/08/99	Espanol	Valencia	3	2	3	0	3	3	3	5	0	-1	7	18
99_00	Ida	2	29/08/99	Real Madrid	Numancia	4	1	6	3	6	2	2	4	4	-2	1	14
99_00	Ida	2	29/08/99	Santander	Barcelona	1	2	3	6	3	4	3	1	0	2	8	2
99_00	Ida	2	29/08/99	Valladolid	Ath Bilbao	1	0	3	3	1	1	1	1	0	0	12	11
99_00	Ida	2	29/08/99	Vallecano	Mallorca	2	1	6	0	4	2	1	4	3	-2	3	19
99_00	Ida	2	29/08/99	Zaragoza	Oviedo	4	0	3	3	4	1	2	4	2	-3	6	15
99_00	Ida	3	11/09/99	Barcelona	Espanol	3	0	9	3	7	3	1	6	6	-3	1	14
99_00	Ida	3	11/09/99	La Coruna	Valladolid	2	0	7	3	6	1	1	3	5	-2	3	13
99_00	Ida	3	11/09/99	Valencia	Alaves	0	2	0	6	3	5	7	5	-4	0	19	8
99_00	Ida	3	12/09/99	Ath Bilbao	Real Madrid	2	2	4	7	3	8	3	4	0	4	11	4
99_00	Ida	3	12/09/99	Ath Madrid	Celta	1	2	0	6	2	4	8	3	-6	1	20	7
99_00	Ida	3	12/09/99	Malaga	Betis	3	0	6	1	5	0	2	4	3	-4	5	18
99_00	Ida	3	12/09/99	Mallorca	Numancia	3	0	3	3	5	2	4	7	1	-5	12	16
99_00	Ida	3	12/09/99	Oviedo	Santander	1	2	3	6	2	5	6	4	-4	1	15	6
99_00	Ida	3	12/09/99	Sevilla	Zaragoza	0	0	2	4	3	4	4	2	-1	2	17	10
99_00	Ida	3	12/09/99	Vallecano	Sociedad	2	1	9	4	6	7	2	5	4	2	2	9
99_00	Ida	4	18/09/99	Alaves	Barcelona	2	1	9	9	7	8	6	3	1	5	3	2
99_00	Ida	4	18/09/99	Betis	Valencia	1	0	4	0	1	3	4	8	-3	-5	14	20
99_00	Ida	4	18/09/99	Espanol	Oviedo	2	1	6	3	5	3	7	8	-2	-5	11	18
99_00	Ida	4	18/09/99	Real Madrid	La Coruna	1	1	8	8	9	7	5	2	4	5	5	4
99_00	Ida	4	18/09/99	Valladolid	Malaga	4	2	6	6	5	7	5	6	0	1	9	8
99_00	Ida	4	19/09/99	Celta	Vallecano	0	1	6	12	4	7	4	2	0	5	10	1
99_00	Ida	4	19/09/99	Numancia	Ath Bilbao	1	1	4	5	3	4	8	4	-5	0	15	13
99_00	Ida	4	19/09/99	Santander	Sevilla	2	2	7	3	7	5	6	6	1	-1	7	17
99_00	Ida	4	19/09/99	Sociedad	Mallorca	2	1	7	3	9	6	6	6	3	0	6	16
99_00	Ida	4	19/09/99	Zaragoza	Ath Madrid	1	1	5	1	5	3	3	9	2	-6	12	19
99_00	Ida	5	25/09/99	Barcelona	Betis	4	1	12	4	12	2	4	8	8	-6	1	18
99_00	Ida	5	25/09/99	Malaga	Real Madrid	1	1	7	9	8	10	7	6	1	4	10	3
99_00	Ida	5	25/09/99	Sociedad	Celta	0	2	7	9	9	6	8	4	1	2	9	4
99_00	Ida	5	25/09/99	Valencia	Valladolid	0	0	1	7	3	5	8	5	-5	0	20	11
99_00	Ida	5	26/09/99	Ath Madrid	Santander	2	0	4	7	5	7	9	8	-4	-1	17	12
99_00	Ida	5	26/09/99	La Coruna	Numancia	0	2	8	7	7	5	4	8	3	-3	7	13
99_00	Ida	5	26/09/99	Mallorca	Ath Bilbao	2	1	6	5	8	5	7	6	1	-1	14	16
99_00	Ida	5	26/09/99	Oviedo	Alaves	1	0	6	9	4	7	8	7	-4	0	15	5
99_00	Ida	5	26/09/99	Sevilla	Espanol	1	2	3	9	6	7	8	8	-2	-1	19	6
99_00	Ida	5	26/09/99	Vallecano	Zaragoza	0	1	12	8	7	6	3	3	4	3	2	8

### 3.3.1. Otras representaciones gráficas

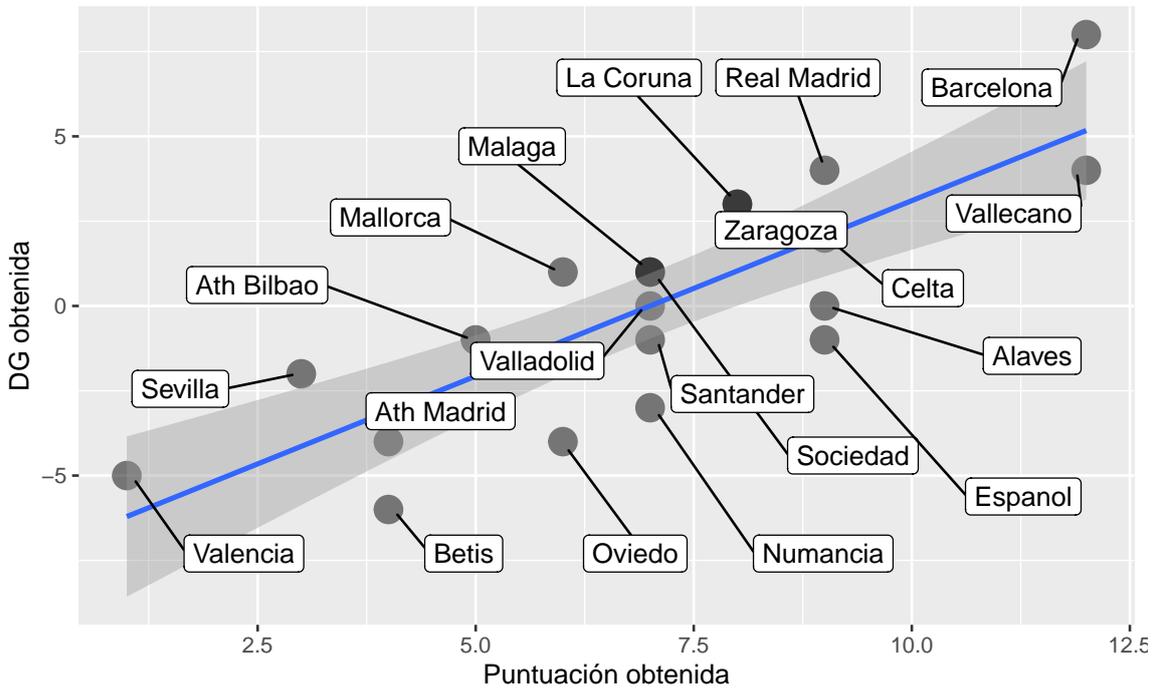
Otra forma de expresar las 5 primeras jornadas de la temporada 1999/2000:

Cuadro 3.2: Clasificaciones correspondientes a las 5 primeras jornadas de la temporada 1999/2000

Temporada 1999/2000					
Posición	Jornada 1	Jornada 2	Jornada 3	Jornada 4	Jornada 5
<b>Campeón</b>					
1	La Coruna	Real Madrid	Barcelona	Vallecano	Barcelona
<b>UCL</b>					
2	Ath Madrid	Barcelona	Vallecano	Barcelona	Vallecano
3	Vallecano	Vallecano	La Coruna	Alaves	Real Madrid
4	Real Madrid	Sociedad	Real Madrid	La Coruna	Celta
<b>UEL</b>					
5	Santander	La Coruna	Malaga	Real Madrid	Alaves
6	Malaga	Zaragoza	Santander	Sociedad	Espanol
<b>Media tabla</b>					
7	Sevilla	Espanol	Celta	Santander	La Coruna
8	Sociedad	Santander	Alaves	Malaga	Zaragoza
9	Mallorca	Celta	Sociedad	Valladolid	Sociedad
10	Valencia	Malaga	Zaragoza	Celta	Malaga
11	Numancia	Ath Bilbao	Ath Bilbao	Espanol	Valladolid
12	Betis	Valladolid	Mallorca	Zaragoza	Santander
13	Ath Bilbao	Alaves	Valladolid	Ath Bilbao	Numancia
14	Celta	Numancia	Espanol	Betis	Mallorca
15	Espanol	Oviedo	Oviedo	Numancia	Oviedo
16	Oviedo	Sevilla	Numancia	Mallorca	Ath Bilbao
17	Valladolid	Betis	Sevilla	Sevilla	Ath Madrid
<b>Descenso</b>					
18	Barcelona	Valencia	Betis	Oviedo	Betis
19	Zaragoza	Mallorca	Valencia	Ath Madrid	Sevilla
20	Alaves	Ath Madrid	Ath Madrid	Valencia	Valencia

Gracias a este tipo de representación gráfica podemos apreciar detalles más complejos que a priori con el Cuadro 3.1 no eran fácilmente observables, como son la evolución del Alavés con estas 5 jornadas, donde comenzó en la primera jornada siendo colista y acabando en la quinta jornada en puestos de UEL. Otro caso a destacar es la del Deportivo de la Coruña, el cual comenzó siendo líder y acabó en la jornada 5 fuera de puestos de zona europea; en media tabla.

Relación Puntuación/DG obtenida  
(Temporada 1999/2000. Jornada 5)



Esta es otra forma representativa de los datos, en la cual hemos mostrado gráficamente la relación de puntos obtenidos con la diferencia de goles lograda en la jornada 5 de la temporada 1999/2000 por cada uno de los equipos. Esto sirve como comprobación básica de la importancia de mantener una diferencia de goles, cuanto más positiva mejor, asegurándose un puesto entre los primeros de la clasificación, como ocurre con el Barcelona en el gráfico. Por el contrario también ocurre lo mismo, donde si se tiene una diferencia de goles cada vez peor, se obtendrá una peor posición en la tabla clasificatoria, como es el caso del Valencia en el gráfico.

# Capítulo 4

## Creación BBDD real

El primero de los 3 bloques de clasificaciones finales a lo largo de las temporadas entre 2009/2010 y 2018/2019 será el conjunto de series de datos que corresponden a las clasificaciones finales reales, es decir, las clasificaciones que finalmente se obtuvieron tras la disputa de todos los encuentros de cada una de las temporadas.

### 4.1. Definición de éxito probabilístico

Conforme a nuestros criterios de predicción y en base a nuestro objetivo final estableceremos nuestra propia definición de éxito probabilístico. Ésta será nuestra forma de medir la fiabilidad de cada uno de nuestros modelos, a lo que llamaremos que hemos obtenido éxito, puesto que lo utilizaremos como comparación entre las posiciones obtenidas de los distintos equipos durante cada temporada. Se tomará de tal forma que se nos indique un porcentaje de fiabilidad para 5 grupos distintos en nuestra clasificación final:

- **Campeón:** el primer grupo para medir nuestro acierto será el del campeón, el cual será 100 % si se acierta o 0 % si se falla.
- **UCL:** este grupo consta de las posiciones correspondientes a los 4 equipos que tras la disputa de las 38 jornadas queden en los 4 primeros puestos, obteniendo como “*premio*” final de la temporada el poder disputar la próxima campaña la Uefa Champions League, considerada la mejor competición de clubes por excelencia. Se obtendrá un 100 % de acierto en caso de haber acertado los 4 mejores combinados de LaLiga Santander al final de la competición doméstica.
- **UEL:** para este grupo hay una cierta particularidad, puesto que a priori participarían en la próxima campaña en la Uefa Europa League los equipos resultantes en los puestos 5º y 6º tras la finalización de la temporada. Sin embargo, hay excepciones como que el campeón de la UEL de cierta temporada, aunque su posición en la tabla clasificatoria se encuentre por debajo del 4º puesto (5º - 20º), tenga plaza automática en la UCL, por lo que podría haber hasta 5 equipos de un país compitiendo en dicha competición. En nuestro caso, al asociarse con excepciones, tomaremos simplemente a los equipos del grupo UEL a aquellos que queden en 5º y 6º lugar en la clasificación.
- **Media tabla:** a este tramo de la clasificación se le conoce como *Media tabla* puesto que abarcan los puestos de la clasificación en el que los equipos que los ocupan simplemente logran el objetivo de la permanencia en LaLiga Santander por una

temporada más; no acaban descendiendo ni jugando en ninguna competición europea. Es el grupo más abultado puesto que ocupa 11 de los 20 puestos, desde la posición 7 hasta la número 17.

- **Descenso:** por último tenemos el grupo del *descenso*, el cual lo ocupan los 3 peores equipos de la temporada tras haberse disputado todas los encuentros. Estos equipos son correspondidos con los puestos 18, 19 y 20 de la clasificación.

**Nota importante:** cuando decimos que se deben acertar  $x$  equipos de un grupo para tener una mayor fiabilidad nos referimos a que dichos equipos se sitúen en el grupo que se está estudiando en ese momento, es decir, no importa que acabe el *Team A* en el puesto 12 en la clasificación de alguno de nuestros modelos y, sin embargo, en la clasificación real de esa misma temporada el *Team A* ha finalizado en la posición 15, puesto que igualmente, en ambos casos el *Team A* se encuentra en el grupo de *Media tabla*. Aquí un ejemplo más sencillo:

	Clasif. Real	Clasif. Modelo
Campeón	Real Madrid	Barcelona
UCL	Barcelona	Real Madrid
UCL	Atlético de Madrid	Atlético de Madrid
UCL	Getafe	Valencia

En este ejemplo tenemos 2 columnas correspondientes a la clasificación real y la clasificación obtenida por un modelo en una determinada temporada. Podemos afirmar que el acierto del grupo *Campeón* es del 0% puesto que no se ha acertado; mientras que el acierto conseguido en el grupo de *UCL* es del 75%, puesto que aunque el *Real Madrid* o el *Atlético de Madrid* no mantengan el mismo puesto en ambas clasificaciones, se dará como éxito ambos equipos puesto que los dos se encuentran en el grupo de *UCL*.

## 4.2. Modificaciones realizadas

A diferencia de la base de datos completa creada en el capítulo 3, ésta es bastante más simple, puesto que no tendremos apenas que operar con los datos en ningún momento al tratarse de obtener únicamente las clasificaciones finales de las temporadas completas de los 10 últimos años. Para ello lo único que hacemos es:

- 1º. Obtener los subconjuntos de la jornada 38 de cada una de las temporadas.
- 2º. Nos quedamos únicamente con las columnas respectivas a *Temporada*, *Local*, *Visitante*, *Ranking local* y *Ranking visitante*.
- 3º. Extraemos de cada temporada la clasificación final en función del orden de los equipos en *Local* y *Visitante*, ya que tendrán su puesto asignado en las columnas *Ranking local* y *Ranking visitante*.
- 4º. Finalmente unimos todas las clasificaciones en una sólo tabla para poder manejar todos los datos necesarios con un sólo archivo.

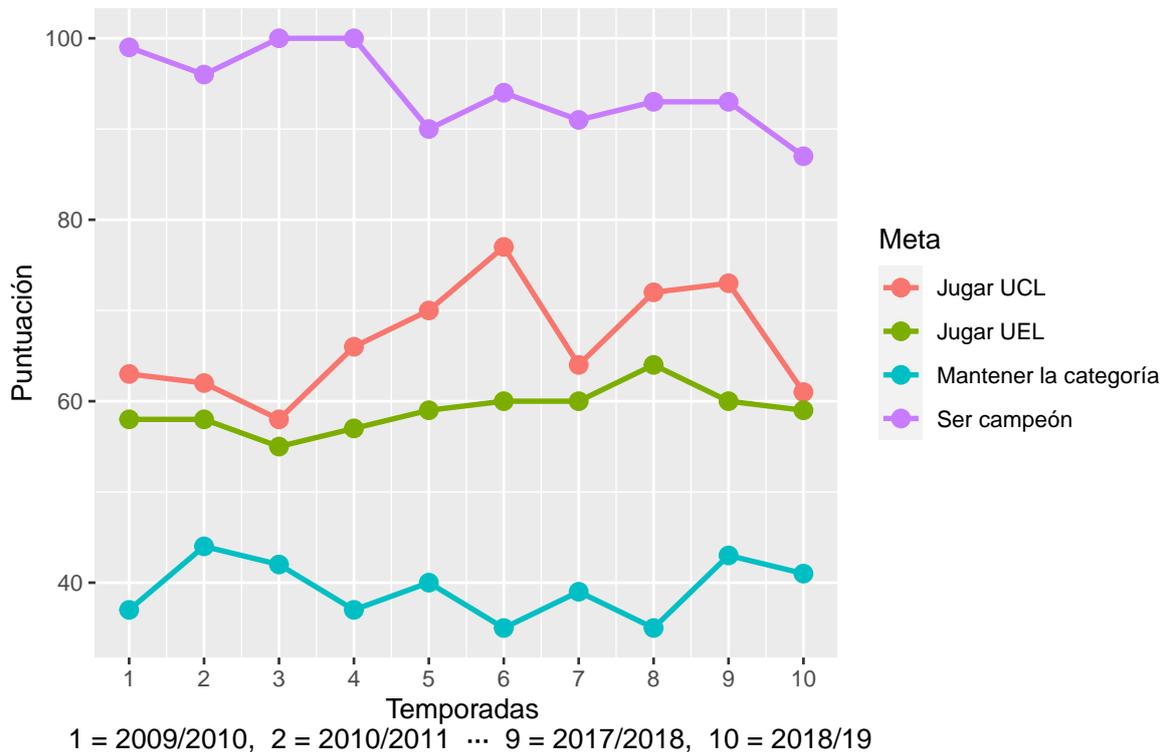
## 4.3. Tabla gráfica

Cuadro 4.1: Clasificaciones finales comprendidas entre las temporadas 2009/2010 y 2018/2019

Posición	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
<b>Campeón</b>										
1	Barcelona	Barcelona	Real Madrid	Barcelona	Ath Madrid	Barcelona	Barcelona	Real Madrid	Barcelona	Barcelona
<b>UCL</b>										
2	Real Madrid	Real Madrid	Barcelona	Real Madrid	Barcelona	Real Madrid	Real Madrid	Barcelona	Ath Madrid	Ath Madrid
3	Valencia	Valencia	Valencia	Ath Madrid	Real Madrid	Ath Madrid	Ath Madrid	Ath Madrid	Real Madrid	Real Madrid
4	Sevilla	Villarreal	Malaga	Sociedad	Ath Bilbao	Valencia	Villarreal	Sevilla	Valencia	Valencia
<b>UEL</b>										
5	Mallorca	Sevilla	Ath Madrid	Valencia	Sevilla	Sevilla	Ath Bilbao	Villarreal	Villarreal	Getafe
6	Getafe	Ath Bilbao	Levante	Malaga	Villarreal	Villarreal	Celta	Sociedad	Betis	Sevilla
<b>Media tabla</b>										
7	Villarreal	Ath Madrid	Osasuna	Betis	Sociedad	Ath Bilbao	Sevilla	Ath Bilbao	Sevilla	Espanol
8	Ath Bilbao	Espanol	Mallorca	Vallecano	Valencia	Celta	Malaga	Espanol	Getafe	Ath Bilbao
9	Ath Madrid	Sp Gijon	Sevilla	Sevilla	Celta	Malaga	Sociedad	Alaves	Eibar	Sociedad
10	La Coruna	Osasuna	Ath Bilbao	Getafe	Levante	Espanol	Betis	Eibar	Girona	Betis
11	Espanol	Malaga	Getafe	Levante	Malaga	Vallecano	Las Palmas	Malaga	Espanol	Alaves
12	Osasuna	Santander	Sociedad	Ath Bilbao	Vallecano	Sociedad	Valencia	Valencia	Sociedad	Eibar
13	Almeria	Zaragoza	Betis	Espanol	Getafe	Elche	Espanol	Celta	Celta	Leganes
14	Zaragoza	Sociedad	Espanol	Valladolid	Espanol	Levante	Eibar	Las Palmas	Alaves	Villarreal
15	Sp Gijon	Levante	Vallecano	Granada	Granada	Getafe	La Coruna	Betis	Levante	Levante
16	Santander	Getafe	Zaragoza	Osasuna	Elche	La Coruna	Granada	La Coruna	Ath Bilbao	Valladolid
17	Malaga	Mallorca	Granada	Celta	Almeria	Granada	Sp Gijon	Leganes	Leganes	Celta
<b>Descenso</b>										
18	Tenerife	La Coruna	Villarreal	Mallorca	Osasuna	Eibar	Vallecano	Sp Gijon	La Coruna	Girona
19	Valladolid	Hercules	Sp Gijon	La Coruna	Valladolid	Almeria	Getafe	Osasuna	Las Palmas	Huesca
20	Xerez	Almeria	Santander	Zaragoza	Betis	Cordoba	Levante	Granada	Malaga	Vallecano

### 4.3.1. Otras formas de representación gráfica

Puntuación necesaria para lograr cada una de las situaciones



Aquí podemos observar otra forma de llevar a cabo una representación gráfica para sacar unas conclusiones algo más difíciles de ver a simple vista, y más aún siendo la comparación de 10 temporadas a la vez. En este caso se ha representado la cantidad de puntos necesarios para poder lograr alguna de las situaciones que hemos definido. Para llevar a cabo una buena visión general, hemos optado por representar los puntos que fueron necesarios en cada temporada para ser campeón, poder jugar UCL la próxima campaña, jugar UEL la próxima campaña y evitar el descenso, o lo que es lo mismo, mantener la categoría.

Notemos algunas observaciones como:

- La puntuación necesaria para ser campeón tiene una ligera tendencia descendente, indicando que cada temporada va habiendo más igualdad entre los distintos equipos.
- La puntuación necesaria para mantener la categoría ha sido generalmente estable, aunque con algunos pequeños repuntes.
- La puntuación requerida para lograr una plaza para jugar UCL ha mantenido una línea ligeramente ascendente, aunque esta última temporada tuvo una caída considerable.
- En la temporada 2014/2015 (punto 6) se estableció una mayor diferencia entre los equipos que disputaban UCL y los que disputaban UEL; sin embargo, esta última campaña han estado casi igualados a puntuación, apenas sin haber diferencia.
- Por último, la tendencia de LaLiga ha sido ir igualándose poco a poco, puesto que hasta la temporada 2013/2014 había una diferencia abismal de puntuación entre los equipos con plaza *Champions* y el campeón; no obstante, desde esa temporada a la actual ha mantenido ligeramente una tendencia de ir aproximándose entre ellos.

# Capítulo 5

## Sistema de puntuación ELO

### 5.1. Introducción

El sistema de puntuación ELO se basa en estimar la probabilidad estadística de las posibles distintas situaciones de un juego, así como las habilidades relativas de algunos deportes como es el ajedrez. Este sistema se ha aplicado tanto a deportes individuales como grupales. Sirve como otra variante para establecer clasificaciones o rankings.

#### 5.1.1. Un poco de historia

Esta metodología fue creada por el físico estadounidense de origen búlgaro Arpad Elo, de ahí el nombre del sistema. Arpad Elo trabajó para la USCF (Federación de Ajedrez de Estados Unidos), donde le asignaron el puesto de presidente del Comité de Puntuación en 1959 y, tan sólo un año más tarde, asignaron de manera oficial en la USCF este sistema de puntuación. 10 años más tarde ya se oficializó por parte de la FIDE (Federación Internacional de Ajedrez) hasta el día de hoy, donde no ha sufrido apenas cambios en el sistema de puntuación o ponderación.

#### 5.1.2. Dónde podemos verlo en la actualidad

El sistema ha sido utilizado por más sectores además del ajedrez a nivel mundial. Uno de los principales casos conocidos es la aplicación telefónica *Tinder*; aplicación social creada en 2012 con el fin de que sus usuarios consigan hacer *Matches* con otros usuarios y poder chatear. Lo que le hace diferente al resto de aplicaciones y le haya hecho obtener tanta fama mundial es que en esta aplicación los usuarios pueden escoger o indicar qué usuarios son o no de su agrado para concretar citas o encuentros posteriormente.

El algoritmo que presenta *Tinder* usa un sistema de puntuación ELO, donde cada usuario tiene una propia puntuación que varía en función de muchos factores, determinando varias estadísticas como son el número de veces que se muestra tu perfil a otros usuarios, qué perfiles se te muestran al abrir la aplicación o a qué perfiles muestran tu usuario. Hasta el año pasado mantuvo este sistema, pero lo cambió debido a que ya lo consideraban algo obsoleto para la finalidad que buscaban [25].

### 5.1.3. ¿Qué deportes utilizan dicha puntuación?

No obstante, no hay que dejar de lado que el punto más fuerte del sistema ELO es el mundo del deporte. Desde 1960 ya se trata de forma oficial este sistema para la USCF y, posteriormente, para la FIDE.

El ajedrez es el deporte que más ha repercutido para el puntaje ELO, pero también hay otros varios casos. La otra principal fuente oficial donde se aplica el ELO es la clasificación mundial de FIFA [18], organización europea de las distintas federaciones de los equipos nacionales de fútbol de todo el mundo. Dicha organización hasta hace pocas temporadas presentaba otro sistema de puntuación completamente distinto al que se aplica ahora, pero mantuvo muchas críticas al haber selecciones nacionales en puestos que la mayoría de seguidores daba por altamente improbable.

Así ocurrió con la selección de Suiza situándose en el 5º puesto en 2017, antes del comienzo del pasado mundial de fútbol de Rusia en 2018 [16]. Como era de esperar, Suiza cayó a las primeras de cambio en los octavos de final ante Suecia, selección que ocupaba el puesto número 18 [15].

Por otro lado, este sistema se ha utilizado para llevar a cabo ciertos estudios en otros deportes, como son la NBA, el fútbol a nivel de clubes, o el tenis. Éste último ha sido muy trabajado, puesto que el sistema que se aplica actualmente para determinar el puesto de cada tenista es el ranking ATP. Mantiene diferencias respecto al ELO, pero se le asemeja bastante al tratarse de un deporte individual y permite con esto procesar comparaciones en ambas clasificaciones sin mucho margen de error. Este sistema, a diferencia del antiguo sistema de clasificación ranking de la FIFA, no presenta muchas críticas al ajustarse mucho más a la realidad.

## 5.2. The “ELO” Package

Aquí comenzamos el primero de los modelos que vamos a tomar para medir su fiabilidad en base a los resultados reales de las pasadas 10 campañas en LaLiga Santander. El programa RStudio forma parte del software libre R, que se basa en el estudio de análisis estadísticos y gráficas. Presenta un alto índice de popularidad en el mundo del análisis de datos, puesto que presenta una curva de aprendizaje relativamente constante y permite la opción de compartir código con otras personas, lo que puede llegar a ser muy útil entre investigadores que están profundizando en sus conocimientos estadísticos. Por ello el software R se encuentra entre los 3 softwares de análisis de datos más utilizados del mundo, junto a Python y MySQL [20].

Dentro de RStudio podemos acceder a cientos de paquetes con funciones de todo tipo, incluyendo el paquete ELO, el cual nos ayudará en este trabajo. A raíz de este paquete estadístico podremos llevar a cabo las predicciones de las pasadas 10 temporadas y su consiguiente fiabilidad en base a los aciertos obtenidos en las mismas. Este paquete tiene instaurado en él el sistema de puntuación que ejecuta la puntuación ELO y, con ello, las fórmulas necesarias para hacer predicciones sobre eventos deportivos. Nos proporciona las probabilidades de éxito de cada equipo y/o empate en caso de poder darse tal caso.

### 5.2.1. Posibilidad de ocurrir un empate en un juego

Ejemplos como son los partidos de fútbol se tendrá en cuenta la posibilidad de acabar en empate el encuentro, puesto que no estamos hablando de predecir un partido que corresponda a una eliminatoria como puede ser un partido de Copa del Rey o de la Uefa Champions League. Esta situación podemos encontrarla también en el ajedrez, donde se pueden producir la llamada situación de *tablas*. Se le denominan *tablas* a aquella partida donde ambos oponentes tras haber disputado al menos 50 jugadas entre los dos sin lograr ninguna captura o sin avanzar ningún peón [8]; a partir de ahí uno de los dos oponentes tiene la opción de solicitar tablas a su adversario. Resaltar que también existen otras situaciones de posibles tablas.

Sin embargo esto no ocurre siempre. Hay deportes en el que el empate no es opción, como es el caso de los deportes de raqueta o el baloncesto. El deporte más conocido de raqueta es el tenis, deporte en el que se pueden enfrentar individualmente o por parejas, pero no viéndose afectado la imposibilidad de ocurrir un empate. Cuando un partido de tenis se encuentra en el último *set* e igualado a 6 en número de *juegos*, los jugadores deberán disputar tantos *juegos* como sean necesarios hasta obtener una diferencia de 2. El caso más excepcional en este sentido fue el partido más largo de la historia del tenis, disputado entre el estadounidense John Isner y el francés Nicolas Mahut, donde en el 5º y último *set* el estadounidense acabó llevándose el gato al agua con un total de 70 *juegos* a 68 por parte del francés [1].

Por último añadir el caso del baloncesto, uno de los deportes más practicados y seguidos del mundo. En este deporte cada partido se constituye de 4 períodos de tiempo, donde si tras finalizar el último en caso de empate se disputa una prórroga de otros 5 minutos de añadido. Este proceso puede llevarse a cabo tantas veces como ambos equipos sigan empatando en dichas prórrogas; en caso de finalizar la prórroga en desigualdad tendríamos a nuestro ganador del encuentro. Es el caso del partido más largo de la historia de la NBA, donde se disputaron un total de 6 prórrogas hasta que finalmente el equipo de Indianapolis se llevase el encuentro frente a los Royals en un marcador final de 75 a 73 [12].

### 5.2.2. Funciones utilizadas y qué opciones útiles nos aporta

El paquete ELO de RStudio presenta muchas alternativas y funciones para llevar a cabo distintas operaciones estadísticas, según nuestros objetivos. Para este caso de la predicción de partidos de fútbol de LaLiga Santander hemos optado por el empleo de estas funciones provenientes de [14]:

- **elo.prob()**: la función *elo.prob()* se utiliza para obtener la probabilidad de ganar un equipo *Team A* a otro equipo *Team B*, dando como resultado un valor numérico comprendido entre 0 y 1 en función a la puntuación ELO de cada uno de los equipos.
- **elo.update()**: la función *elo.update()* se le asigna como objetivo calcular qué variación obtendrán *Team A* y *Team B* en su puntuación actual ELO tras haber disputado un partido entre ambos conjuntos, sumando o restando en sus puntuaciones en función de si han conseguido la victoria o han sido derrotados.

- **elo.calc()**: la función *elo.calc()* nos aporta la nueva puntuación ELO de ambos equipos tras la finalización del encuentro disputado entre ambos. Dichas variaciones respecto a las puntuaciones anteriores son las calculadas previamente por la función *elo.update()*.
- **elo.run()**: la función *elo.run()* será la que nos aporte nuestro modelo de ajuste para nuestra serie de datos, calculando las actualizaciones de puntuación ELO de los mismos. Esta función presenta variantes para distintos casos como es disputar un partido siempre contra un oponente de puntuación ELO fija, es decir, nunca varía. En nuestro caso van a ser equipos con variaciones constantes en sus puntuaciones, por lo que hemos utilizado la función *elo.run()* en el caso de tener dos variables con puntuaciones ELO cambiantes.
- **score()**: la función *score()* se encuentra en el interior de la función *elo.run()* y nos ayudará a calcular a los ganadores de los partidos de nuestra serie de datos sobre la marcha (1 = victoria, 0.5 = empate, 0 = derrota).
- **k()**: la función *k()* se encuentra en el interior de la función *elo.run()* y nos aportará dificultar un poco más las actualizaciones de las puntuaciones ELO, como en nuestro caso, tomando el logaritmo del margen de victoria como parte de la actualización ELO.
- **final.elos()**: la función *final.elos()* nos dotará de las puntuaciones finales de cada equipo tras la disputa de cada jornada.
- **rank()**: la función *rank()* nos facilitará el orden de posición de los distintos equipos en base a la puntuación calculada previamente con la función *final.elos()*.
- **predict()**: la función *predict()* sirve para la predicción de un cierto partido en concreto entre *Team A* y *Team B* en base a las puntuaciones ELO obtenidas según los resultados conseguidos previamente en nuestra serie de datos.

### 5.3. Creación BBDD puntuación ELO

Como segundo principal pilar de nuestro estudio tenemos al modelo de predicción regido por el sistema de puntuación ELO, el cual nos proporcionará una estimación con una cierta fiabilidad de acierto con el que estimar el próximo campeón de LaLiga Santander tras el parón causado por el COVID-19. Para poder llegar a esas conclusiones finales que queremos lograr debemos obtener una base de datos adaptada a las variables necesarias para la posible utilización del propio método estadístico. Para ello partimos de la base de datos creada inicialmente, la cual tomaremos como referencia estadística para la creación de la misma.

### 5.3.1. Modificaciones realizadas respecto a BBDD inicial

Al igual que la base de datos creada en el capítulo 4 para la obtención de las clasificaciones reales logradas en las temporadas entre 2009/2010 y 2018/2019, en este caso también se ha llevado el mismo procedimiento creando una propia base de datos tomando como guía la obtenida en el capítulo 3 de forma inicial y más completa que el resto. En este caso será algo más complejo que la obtención de la base de datos de las clasificaciones reales, puesto que se creará alguna que otra nueva variable necesaria para el sistema de puntuación ELO.

Las variables utilizadas para nuestra base de datos son las columnas correspondientes a los datos que hacen referencia a *Temporada*, *Jornada*, *Local*, *Visitante*, *Gol local*, *Gol visitante* y una nueva variable que denominaremos por *Gana.A*.

Esta nueva columna que hemos denotado como *Gana.A* hace referencia a si el equipo situado en la variable *Local* ha anotado más, los mismos o menos goles que el equipo referente a la columna *Visitante*, es decir, nos proporcionaría la información sobre si el equipo *Local* ha ganado, empatado o perdido el encuentro frente al equipo *Visitante*. Los datos quedarán almacenados de tal forma que siempre hará referencia al resultado obtenido por el equipo *Local* y nunca el *Visitante*, tomando 3 posibles valores en dicha variable:

- 1 en caso de que el equipo *Local* consiga doblegar al equipo *Visitante*.
- 0.5 sería para las situaciones donde el encuentro haya finalizado por igualdad de goles anotados.
- 0 en la única opción donde el equipo *Visitante* se lleva la victoria en el encuentro aún jugando “*fuera de casa*”.

### 5.3.2. Tratamiento y depuración de datos

Una vez obtenemos las correspondientes variables citadas anteriormente debemos tener en cuenta un detalle principal para la creación de nuestra base de datos y poder aplicarla en el método predictivo del sistema de puntuación ELO, y es que debemos separar nuestros datos en dos subconjuntos.

El primer subgrupo que definiremos serán aquellos resultados obtenidos desde la jornada 1 hasta las 27 inclusive para cada una de las temporadas desde la 2009/2010 para lograr así a encontrarnos en la misma situación que la que nos encontramos en la temporada actual y poder lograr predicciones que se puedan comparar de la misma manera y obtener conclusiones reales acerca de su probabilidad y fiabilidad de acierto.

El otro subconjunto que definiremos a partir de los datos provenientes de las columnas nombradas en el apartado anterior será únicamente el de las columnas *Local* y *Visitante* referentes a las últimas 11 jornadas de cada una de las 10 temporadas de nuestra base de datos (jornadas pertenecientes entre la 28 y 38, ambas inclusive), aportándonos única y exclusivamente los encuentros que quedarían por disputarse, y así lograr la misma situación a la actual de igual forma que en anterior subconjunto.

### 5.3.3. Funcionamiento del método

Este modelo de predicción basado en el sistema de puntuación ELO parte de una base en el que todos los equipos de la competición comienzan con una puntuación asignada base de 1500 puntos ELO. Esta puntuación irá tomando variaciones, positivas o negativas, aunque también puede darse el caso de no variar tras la disputa de un encuentro. Estas variaciones son calculadas en función a la puntuación correspondiente a cada equipo del encuentro obtenidas de la jornada anterior.

Serán más o menos positivas en función a si un equipo *Team A*, de una puntuación ELO baja, logra batir a otro equipo *Team B* a priori superior al tener una puntuación ELO bastante alta; de igual forma el *Team B* se vería bastante perjudicado puesto que tomaría un descenso considerable en su puntuación ELO actual. Esta nueva puntuación lograda por cada equipo, sin tener en cuenta si es mejor o peor que la anterior, es la que se tomará en cuenta para la estimación del próximo encuentro contra otro equipo distinto en la siguiente jornada.

Sin embargo, ya sea bien en una disputa entre dos equipos con puntuación baja o dos equipos con puntuación alta, mientras que haya una diferencia mínima entre ambos conjuntos en su puntuación ELO, éstos no se verán muy afectados, ni positivamente ni negativamente sea cual sea el resultado.

Las diferencias de puntuaciones entre los equipos que se encuentren en las posiciones altas de la tabla clasificatoria y los equipos de peor puntuación serán observables con un transcurso mínimo de varios encuentros para poder catalogarlos a cada uno de una manera correcta y fiable. Para cada encuentro se establece una probabilidad de que el *Team A* bata al *Team B* y viceversa, que será la que tomaremos como el resultado esperado en la columna predefinida anteriormente como *Gana.A*.

La asignación del valor de cada registro en la columna *Gana.A* será 1, 0 ó 0.5, lo cual se verá afectada por dicha probabilidad, aunque no de ninguna manera. En caso de que la probabilidad de que el *Team A* consiga la victoria esté comprendida entre 0.56 y 1 se le asignará un 1 en la columna *Gana.A*; si la probabilidad se sitúa entre 0.45 y 0.55 daremos el resultado final del encuentro como un empate y, por tanto, se le asignará un 0.5 a la columna *Gana.A*. Si la probabilidad de que el *Team A* consiga la victoria es inferior a 0.45 se le reconocerá como derrota del encuentro y aparecerá un 0 en la variable *Gana.A*. En el siguiente cuadro queda reflejado lo explicado:

Cuadro 5.1: Asignación valor *Gana.A* en base a la probabilidad de ganar un encuentro

	Probabilidad gana local	Valor <i>Gana.A</i>
Gana Local	$p > 0.55$	1
Empate	$0.45 < p < 0.55$	0.5
Gana Visitante	$p < 0.45$	0

### 5.3.4. Tablas gráficas

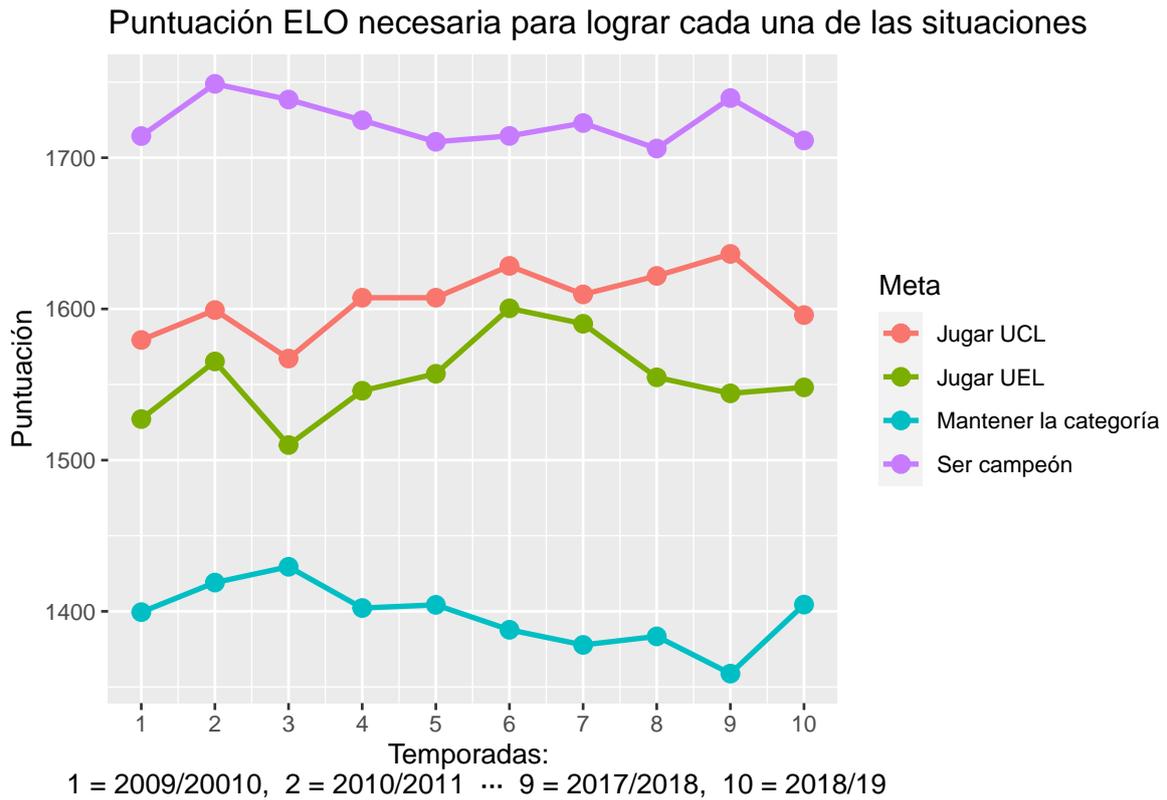
En este apartado veremos algunas formas de representar los datos correspondientes al sistema de puntuación ELO, como son la cabecera de la base de datos, las tablas clasificatorias finales obtenidas por el modelo o un gráfico mostrando visualmente la evolución de la puntuación obtenida por los equipos por el paso de las temporadas.

Cuadro 5.2: Cabecera base de datos sistema de puntuación ELO

Temporada	Jornada	Local	Visitante	Gol_Loc	Gol_Vis	Gana.A
09_10	1	Real Madrid	La Coruna	3	2	1.0
09_10	1	Zaragoza	Tenerife	1	0	1.0
09_10	1	Almeria	Valladolid	0	0	0.5
09_10	1	Ath Bilbao	Espanol	1	0	1.0
09_10	1	Malaga	Ath Madrid	3	0	1.0
09_10	1	Mallorca	Xerez	2	0	1.0
09_10	1	Osasuna	Villarreal	1	1	0.5
09_10	1	Santander	Getafe	1	4	0.0
09_10	1	Valencia	Sevilla	2	0	1.0
09_10	1	Barcelona	Sp Gijon	3	0	1.0
09_10	2	Ath Madrid	Santander	1	1	0.5
09_10	2	Espanol	Real Madrid	0	3	0.0
09_10	2	Getafe	Barcelona	0	2	0.0
09_10	2	Sevilla	Zaragoza	4	1	1.0
09_10	2	La Coruna	Malaga	1	0	1.0
09_10	2	Sp Gijon	Almeria	1	0	1.0
09_10	2	Tenerife	Osasuna	2	1	1.0
09_10	2	Valladolid	Valencia	2	4	0.0
09_10	2	Villarreal	Mallorca	1	1	0.5
09_10	2	Xerez	Ath Bilbao	0	1	0.0
09_10	3	Barcelona	Ath Madrid	5	2	1.0
09_10	3	La Coruna	Espanol	2	3	0.0
09_10	3	Malaga	Santander	1	2	0.0
09_10	3	Mallorca	Tenerife	4	0	1.0
09_10	3	Osasuna	Sevilla	0	2	0.0
09_10	3	Almeria	Getafe	1	0	1.0
09_10	3	Ath Bilbao	Villarreal	3	2	1.0
09_10	3	Real Madrid	Xerez	5	0	1.0
09_10	3	Valencia	Sp Gijon	2	2	0.5
09_10	3	Zaragoza	Valladolid	1	2	0.0
09_10	4	Santander	Barcelona	1	4	0.0
09_10	4	Sevilla	Mallorca	2	0	1.0
09_10	4	Ath Madrid	Almeria	2	2	0.5
09_10	4	Espanol	Malaga	2	1	1.0
09_10	4	Getafe	Valencia	3	1	1.0
09_10	4	Tenerife	Ath Bilbao	1	0	1.0
09_10	4	Valladolid	Osasuna	1	2	0.0
09_10	4	Villarreal	Real Madrid	0	2	0.0
09_10	4	Xerez	La Coruna	0	3	0.0
09_10	4	Sp Gijon	Zaragoza	1	1	0.5

Cuadro 5.3: Clasificaciones finales obtenidas a través del sistema de puntuación ELO

Posición	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
<b>Campeón</b>										
1	Barcelona	Barcelona	Real Madrid	Barcelona	Real Madrid	Barcelona	Barcelona	Barcelona	Barcelona	Barcelona
<b>UCL</b>										
2	Real Madrid	Real Madrid	Barcelona	Real Madrid	Barcelona	Real Madrid	Real Madrid	Real Madrid	Ath Madrid	Ath Madrid
3	Valencia	Valencia	Valencia	Ath Madrid	Real Madrid	Real Madrid				
4	Sevilla	Villarreal	Ath Bilbao	Sociedad	Ath Bilbao	Valencia	Villarreal	Sevilla	Valencia	Getafe
<b>UEL</b>										
5	Mallorca	Ath Bilbao	Ath Madrid	Malaga	Sociedad	Villarreal	Sevilla	Villarreal	Girona	Sevilla
6	Ath Bilbao	Ath Madrid	Malaga	Betis	Sevilla	Sevilla	Ath Bilbao	Sociedad	Getafe	Valencia
<b>Media tabla</b>										
7	Getafe	Espanol	Espanol	Valencia	Villarreal	Sociedad	Malaga	Espanol	Celta	Eibar
8	Ath Madrid	Sevilla	Osasuna	Valladolid	Valencia	Malaga	Eibar	Ath Bilbao	Villarreal	Betis
9	Villarreal	Sociedad	Levante	Vallecano	Espanol	Espanol	Sociedad	Eibar	Eibar	Alaves
10	Sp Gijon	Sp Gijon	Sociedad	Getafe	Levante	Ath Bilbao	Celta	Alaves	Sevilla	Sociedad
11	Malaga	Mallorca	Sevilla	Sevilla	Celta	Celta	Valencia	Celta	Sociedad	Ath Bilbao
12	La Coruna	Santander	Mallorca	Levante	Malaga	Vallecano	La Coruna	Las Palmas	Ath Bilbao	Leganes
13	Almeria	Levante	Vallecano	Espanol	Granada	Getafe	Betis	Valencia	Betis	Villarreal
14	Santander	Osasuna	Betis	Osasuna	Valladolid	Eibar	Las Palmas	La Coruna	Espanol	Girona
15	Osasuna	La Coruna	Getafe	Ath Bilbao	Osasuna	La Coruna	Granada	Malaga	Leganes	Espanol
16	Espanol	Getafe	Granada	Zaragoza	Elche	Almeria	Vallecano	Betis	Alaves	Levante
17	Valladolid	Zaragoza	Santander	Granada	Almeria	Levante	Sp Gijon	Leganes	Levante	Celta
<b>Descenso</b>										
18	Zaragoza	Hercules	Villarreal	Celta	Getafe	Elche	Espanol	Granada	La Coruna	Huesca
19	Tenerife	Almeria	Sp Gijon	Mallorca	Vallecano	Cordoba	Levante	Sp Gijon	Malaga	Vallecano
20	Xerez	Malaga	Zaragoza	La Coruna	Betis	Granada	Getafe	Osasuna	Las Palmas	Valladolid



Aquí tenemos otro tipo de representación gráfica de los datos, siendo ya utilizada en el capítulo 4. En esta ocasión se basa en la puntuación obtenida por el sistema de puntuación ELO, la cual *a priori* no debe mantener los mismos pasos que la puntuación regida por LaLiga actualmente, pero sí debe observarse una similitud en la tendencia de cada una de las situaciones. Podemos notar algunas observaciones en concreto como:

- La puntuación necesaria para ser campeón, a diferencia de la puntuación normal, no presenta una tendencia negativa; más bien constante con algún que otro repunte.
- La puntuación requerida para poder mantener la categoría sí mantiene algunas diferencias respecto a la puntuación normal, puesto que en este caso se aprecia una ligera tendencia negativa. Esto nos quiere decir que ha habido una mayor diferencia entre los equipos de posiciones altas de la tabla clasificatoria con los equipos que ocupan las posiciones bajas, puesto que los peores equipos logran cada temporada una puntuación algo peor que la anterior, a excepción de esta última temporada pasada.

# Capítulo 6

## Modelo de predicción

### 6.1. Introducción

Este presente capítulo sirve para dar por finalizado el apartado de todos los desarrollos y ejecuciones de nuestros modelos de predicción, además de la creación de las distintas bases de datos adaptadas a cada uno de los tres grandes pilares de nuestro trabajo: las clasificaciones reales, las clasificaciones obtenidas a través del sistema de puntuación ELO y ahora por último las clasificaciones obtenidas por el modelo que vamos a crear a continuación.

### 6.2. Objetivo principal

Para ello, procederemos a un modelo de predicción creado por nosotros mismos y basado en los resultados de los encuentros disputados desde la temporada 1999/2000 hasta la actual. Como principal objetivo a marcar es conseguir un porcentaje de acierto en las clasificaciones finales lo más próximo posible al nivel de fiabilidad que presente el sistema de puntuación ELO, el cual se espera considerablemente alto.

Una vez se obtengan los resultados finales procederemos a la representación gráfica de los datos para tener una mejor visualización sobre nuestro modelo. Se mostrarán gráficamente los niveles de acierto, tanto por los grupos de la clasificación como por temporadas de forma individual. Una vez así podremos realizar comparaciones mucho más vistosas junto al sistema de puntuación ELO.

### 6.3. Descripción del modelo

Hemos querido lograr un modelo fiable con el que poder tener un índice de acierto considerable y suela funcionar. Para ello hemos basado nuestro modelo en los resultados y estadísticas de todos los encuentros disputados por cada uno de los equipos pertenecientes a LaLiga Santander para cada una de las temporadas, desde la 2009/2010 hasta la actual.

Estamos ante un modelo de predicción clasificatorio multinomial, el cual está basado en la utilización de herramientas de los conocidos árboles de decisión (Random Forest). De esta forma se generan tantos árboles como se le indique (en nuestro caso hemos generado 1000 árboles), obteniendo los mejores resultados de los mismos y efectuando un promedio sobre ellos, logrando de esta forma reducir la varianza final.

Notar que, como hemos aclarado en el Capítulo 4, la definición de éxito que hemos asignado se corresponde con el grupo final al que pertenecerá cada equipo a final de temporada, puesto que el logro es el mismo aunque haya variaciones en las posiciones de los equipos de un mismo grupo clasificatorio. Por ello, hemos realizado el método *Random Forest* de forma que nos clasifiquen a cada uno de los equipos en los distintos subconjuntos predefinidos anteriormente.

Gráficamente, podemos apreciar mediante el siguiente esquema cómo nuestro modelo quedará constituido por las siguientes secciones o variables predictoras, siendo éstas las que nos ayudará e indicará cómo clasificar a cada equipo en cada uno de los 5 estratos de nuestra variable respuesta.

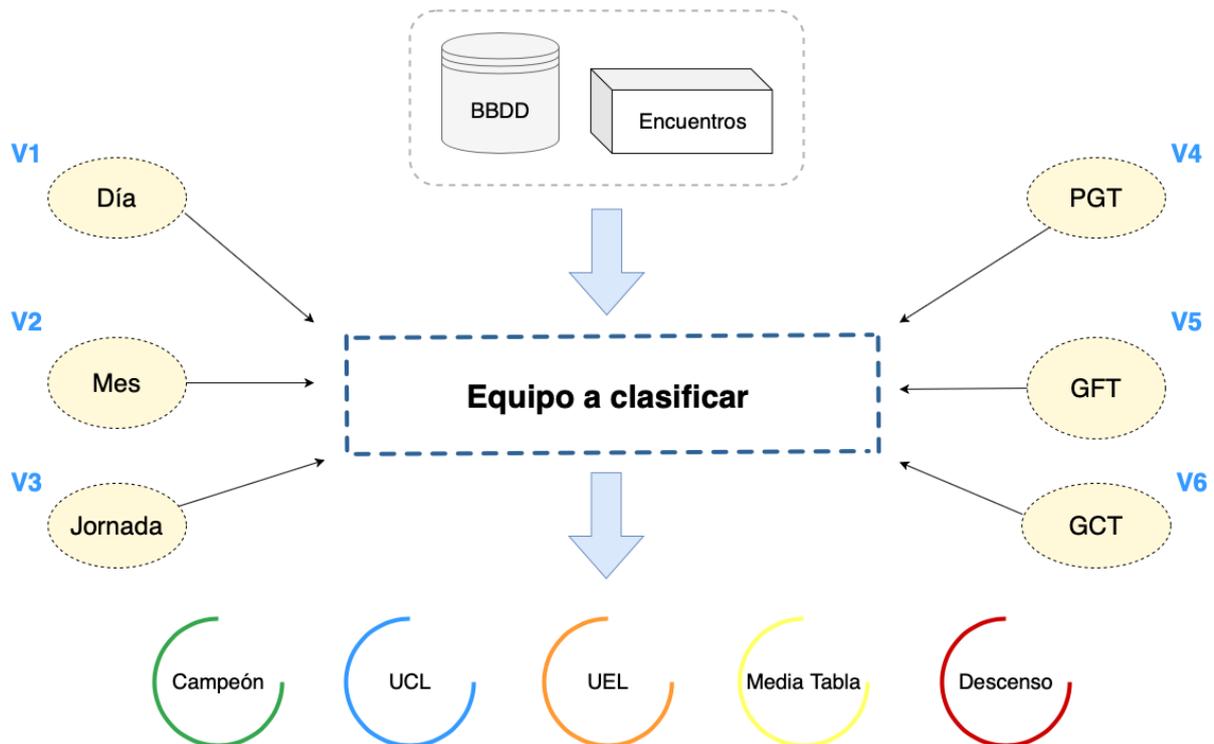


Figura 6.1: Esquema gráfico para clasificar a cada equipo en base a nuestro modelo

### 6.3.1. Variables a tomar en cada partido

Para dilucidar algo más el modelo, aquí se explican tanto cada una de las variables predictoras como la variable respuesta de clasificación multinomial. Nos hemos basado en los datos a nivel temporal de cuándo se disputa cada encuentro y principalmente en algunas estadísticas acumulativas de los equipos durante la temporada que estén disputando:

- **Día:** Esta variable es la primera de las tres que tomaremos como variables predictoras basadas en la cronología de todos los encuentros. Nos aporta el día de la semana en el que se disputó cada partido. No es algo que deba dejarse pasar, puesto que en función a los días de los partidos que se le asignen a cada equipo tendrán más o menos tiempo para preparar el próximo encuentro y descansar lo necesario para llegar óptimos a la próxima disputa.
- **Mes:** Como segunda variable cronológica tenemos el mes en el que el partido se disputó, siendo un punto a considerar. De esta forma podemos clasificar a los equipos en función a su rendimiento físico puesto que en los últimos meses de competición, cuando ya los jugadores arrastran muchos partidos consecutivos, puede variar su eficiencia a la hora de lograr puntos en la clasificación, ya sea para bien o para mal.
- **Jornada:** Al igual que las dos anteriores, esta variable nos ayuda a ver la evolución de los equipos durante la temporada con el transcurso de los encuentros disputados. De esta manera podemos ver qué equipos aumentan su promedio de puntos logrados en las jornadas finales de la temporada (cuando se deciden los puestos) y cuáles no dan la talla en estas situaciones.
- **PGT:** En el ámbito estadístico hemos optado por variables sencillas pero a su vez efectivas para poder clasificar a los equipos. La primera variable es la de *Partidos Ganados Totales*, la cual muestra cuántos partidos ha ganado cada equipo hasta la jornada del encuentro en cuestión. De esta forma podemos obtener detalles como la evolución en victorias de cada equipo o si un equipo se encuentra en racha.
- **GFT:** Una vez vamos decantando el grupo de un equipo en función a sus victorias, podemos entrar en más detalle como es el caso de los *Goles a Favor Totales*, la cual muestra el número de goles anotados de un equipo durante la temporada, haciendo indicar o mostrando cuán fuerte es ese equipo a nivel ofensivo y aspirando a cotas clasificatorias más altas.
- **GCT:** Por último tenemos la variable *Goles en Contra Totales*, que nos servirá como último criterio de clasificación entre los equipos. De esta manera podemos dilucidar qué equipos son más fuertes defensivamente hablando, pudiendo alcanzar mayores logros clasificatorios, puesto que en el fútbol no es sólo anotar goles, sino también evitarlos.

## 6.4. Creación BBDD

Para finalizar, como tercer y último pilar fundamental de nuestro trabajo, tenemos el modelo de predicción clasificatorio multinomial basado en los árboles de decisión. Dicho modelo lo tomaremos como referencia para ser comparado junto a los datos reales para ver su índice de acierto, además de ser comparado conjuntamente con el modelo de predicción basado en el sistema de puntuación ELO.

Para ello vamos a pasar a la creación de una base de datos propia para poder llevarlo a cabo, basada principalmente en la base de datos inicial correspondiente al Capítulo 3.

### 6.4.1. Modificaciones realizadas respecto a BBDD inicial

El proceso de creación de la base de datos referente a nuestro modelo de predicción ha sido generalmente sencilla, puesto que todas las variables se han creado a partir de las pertenecientes a la base de datos inicial sin coste añadido alguno.

**Nota importante:** Para cada una de las columnas de esta base de datos es doble, es decir, lo que corresponde a un único registro en la base de datos inicial explicada en el Capítulo 3 implica haber dos registros en esta nueva tabla, uno por cada equipo del encuentro en cuestión.

Una vez aclarado esto, he aquí la explicación de cómo se han obtenido dichas columnas:

- Variables predictoras:
  - **Temporada:** corresponde a la columna *Temporada* de la tabla inicial.
  - **Día:** hemos extraído parte de la información de la columna *Fecha* de la base de datos inicial en forma de día de la semana.
  - **Mes:** esta variable corresponde a la otra mitad de información correspondiente a *Fecha*, devolviendo los datos en forma de carácter desde *enero* hasta *diciembre*.
  - **Jornada:** nos proporciona la información equivalente a la columna *Jornada* de la tabla inicial.
  - **Equipo:** columna que incluye los datos de todos los equipos de las columnas *Local* y *Visitante* de la base de datos inicial.
  - **PGT:** variable explicada en el paso anterior que implica el número de *Partidos Ganados Totales*. Columna creada a partir de contabilizar aquellos partidos en los que el *Team A* ha anotado más goles que el *Team B*.
  - **GFT:** variable explicada en el paso anterior que implica el número de *Goles Anotados Totales*. Columna creada a partir de contabilizar todos los goles que va logrando un equipo a lo largo de la temporada.
  - **GCT:** variable explicada en el paso anterior que implica el número de *Goles en Contra Totales*. Columna creada a partir de contabilizar todos los goles que va encajando un equipo a lo largo de la temporada.
- Variable respuesta:
  - **Grupo:** esta última columna de la base de datos de nuestro modelo es creada en base al grupo clasificatorio al que pertenece cada equipo durante todos los encuentros que disputa. Se crea dicha clasificación de equipos en base a las columnas *Rank\_Loc* y *Rank\_Vis*, donde un equipo puede pertenecer a uno de los cinco grupos finales representados en la Figura 6.1.

Aquí presentamos la tabla resultante:

Cuadro 6.1: Cabecera base de datos modelo de predicción

Temporada	Dia	Mes	Jornada	Equipo	PGT	GFT	GCT	Grupo
1999-00	4	8	1	Athletic	1	1	0	UEL
1999-00	4	8	1	Mallorca	0	1	2	Media Tabla
1999-00	4	8	1	Valencia	0	1	2	Media Tabla
1999-00	4	8	1	At.Madrid	0	0	2	Descenso
1999-00	4	8	1	Barcelona	1	2	0	UCL
1999-00	4	8	1	D.Coruña	1	4	1	Campeón
1999-00	4	8	1	Malaga	1	1	0	Media Tabla
1999-00	4	8	1	Numancia	1	1	0	Media Tabla
1999-00	4	8	1	Oviedo	1	1	0	Media Tabla
1999-00	4	8	1	Sevilla	0	2	2	Media Tabla
1999-00	4	8	2	Sociedad	1	6	3	UCL
1999-00	4	8	2	Betis	0	0	1	Media Tabla
1999-00	4	8	2	Celta	1	2	2	Media Tabla
1999-00	4	8	2	Alaves	1	3	5	Media Tabla
1999-00	4	8	2	Español	1	3	3	Media Tabla
1999-00	4	8	2	Racing	1	3	3	Media Tabla
1999-00	4	8	2	Rayo	2	4	1	UCL
1999-00	4	8	2	Madrid	2	6	2	Campeón
1999-00	4	8	2	Valladolid	1	1	1	Media Tabla
1999-00	4	8	2	Zaragoza	1	4	2	UEL
1999-00	7	9	3	Barcelona	3	7	1	Campeón
1999-00	7	9	3	D.Coruña	2	6	1	UCL
1999-00	7	9	3	Valencia	0	3	7	Descenso
1999-00	7	9	3	Athletic	1	3	3	Media Tabla
1999-00	7	9	3	At.Madrid	0	2	8	Descenso
1999-00	7	9	3	Malaga	2	5	2	UEL
1999-00	7	9	3	Mallorca	1	5	4	Media Tabla
1999-00	7	9	3	Rayo	3	6	2	UCL
1999-00	7	9	3	Oviedo	1	2	6	Media Tabla
1999-00	7	9	3	Sevilla	0	3	4	Media Tabla
1999-00	7	9	4	Betis	1	1	4	Media Tabla
1999-00	7	9	4	Alaves	3	7	6	UCL
1999-00	7	9	4	Español	2	5	7	Media Tabla
1999-00	7	9	4	Madrid	2	9	5	UEL
1999-00	7	9	4	Valladolid	2	5	5	Media Tabla
1999-00	7	9	4	Celta	2	4	4	Media Tabla
1999-00	7	9	4	Numancia	1	3	8	Media Tabla
1999-00	7	9	4	Racing	2	7	6	Media Tabla
1999-00	7	9	4	Sociedad	2	9	6	UEL
1999-00	7	9	4	Zaragoza	1	5	3	Media Tabla

### 6.4.2. Funcionamiento del modelo

Una vez vistas las variables que componen la base de datos de nuestro modelo, así como la diferenciación entre las variables predictoras y la variable respuesta, ahora pasamos al aspecto del procedimiento que hemos ido realizando; el funcionamiento del modelo.

Partiendo de la base de datos explicada y mostrada anteriormente, hemos dividido el conjunto de datos mediante la función `initial_split()`, función que indicará a la tabla de datos en qué estratos se dividen (son cada uno de los niveles de la columna *Grupo*). Una vez ahí aplicamos las funciones `training()` y `testing()`, con las que crearemos los conjuntos entrenamiento y test respectivamente.

Tomando estos datos pasaremos al siguiente paso. Definimos nuestro tipo de modelo con la función `rand_forest()` y el workflow correspondiente con las funciones `workflow()` y `finalize_workflow()`. Para ajustar los datos hemos utilizado las funciones `fit()` y `last_fit()`.

Por último finalizamos con la obtención de nuestra *matriz de confusión*, donde se presentarán la clasificación que se le ha asignado a cada equipo, mostrando si se ha clasificado o no correctamente. A su vez, podemos comprobar la *accuracy* del modelo mediante el total de la diagonal / total de observaciones de la matriz. Aquí podemos ver la resultante:

Cuadro 6.2: Matriz de confusión obtenida

X	Campeón	Descenso	Media.Tabla	UCL	UEL
Campeón	146	0	0	17	0
Descenso	0	506	70	0	0
Media Tabla	0	101	2027	11	94
UCL	44	0	9	486	81
UEL	0	0	66	71	205

En este caso la *accuracy* obtenida sería de 0.8566, valor más que aceptable.

# Capítulo 7

## Comparativa: resultados 2 a 2

Para este capítulo nos basaremos en hacer comparaciones entre las distintas clasificaciones obtenidas entre las temporadas 2009/2010 y 2018/2019 de LaLiga Santander, tomando siempre de referencia las clasificaciones obtenidas como reales, puesto que son los verdaderos resultados finales que se obtuvieron. El porcentaje de acierto lo tomaremos de dos formas:

- *Por grupos*: la primera forma que tomaremos para medir la fiabilidad del modelo será por grupos, es decir, cada uno de los subconjuntos de la tabla clasificatoria tendrá su propio porcentaje de acierto.
- *Por temporada*: éste servirá como principal dato final que se obtenga sobre la fiabilidad de un modelo en cada temporada, que será el obtenido tras ponderarlo de forma adecuada entre los porcentajes logrados previamente por cada uno de los grupos.

En la comparación referente al apartado 7.3. que veremos más adelante se comparará los modelos del sistema de puntuación ELO y nuestro propio modelo de predicción, no habiendo ninguno 100% fiable con el que tomar referencia, como ocurre con las clasificaciones reales. Sin embargo, tendremos los porcentajes correspondientes de acierto de cada uno de los modelos, pudiendo decantarnos por uno de los dos al presentar un mayor índice de fiabilidad.

### 7.1. Real - ELO

En esta primera comparativa a realizar veremos cuán bueno es el modelo de predicción basado en el sistema de puntuación ELO, donde mostraremos para cada grupo de la tabla clasificatoria y para el total el porcentaje de acierto obtenido.

#### 7.1.1. Tabla gráfica

Aquí presentamos una tabla donde se muestran para cada una de las temporadas las clasificaciones reales y las obtenidas por el sistema de puntuación ELO:

Cuadro 7.1: Clasificaciones reales - puntuación ELO desde la temporada 2009/10 hasta la 2013/14

Posición	2009/10		2010/11		2011/12		2012/13		2013/14	
	Real	ELO								
<b>Campeón</b>										
1	Barcelona	Barcelona	Barcelona	Barcelona	Real Madrid	Real Madrid	Barcelona	Barcelona	Ath Madrid	Real Madrid
<b>UCL</b>										
2	Real Madrid	Real Madrid	Real Madrid	Real Madrid	Barcelona	Barcelona	Real Madrid	Real Madrid	Barcelona	Barcelona
3	Valencia	Valencia	Valencia	Valencia	Valencia	Valencia	Ath Madrid	Ath Madrid	Real Madrid	Ath Madrid
4	Sevilla	Sevilla	Villarreal	Villarreal	Malaga	Ath Bilbao	Sociedad	Sociedad	Ath Bilbao	Ath Bilbao
<b>UEL</b>										
5	Mallorca	Mallorca	Sevilla	Ath Bilbao	Ath Madrid	Ath Madrid	Valencia	Malaga	Sevilla	Sociedad
6	Getafe	Ath Bilbao	Ath Bilbao	Ath Madrid	Levante	Malaga	Malaga	Betis	Villarreal	Sevilla
<b>Media tabla</b>										
7	Villarreal	Getafe	Ath Madrid	Espanol	Osasuna	Espanol	Betis	Valencia	Sociedad	Villarreal
8	Ath Bilbao	Ath Madrid	Espanol	Sevilla	Mallorca	Osasuna	Vallecano	Valladolid	Valencia	Valencia
9	Ath Madrid	Villarreal	Sp Gijon	Sociedad	Sevilla	Levante	Sevilla	Vallecano	Celta	Espanol
10	La Coruna	Sp Gijon	Osasuna	Sp Gijon	Ath Bilbao	Sociedad	Getafe	Getafe	Levante	Levante
11	Espanol	Malaga	Malaga	Mallorca	Getafe	Sevilla	Levante	Sevilla	Malaga	Celta
12	Osasuna	La Coruna	Santander	Santander	Sociedad	Mallorca	Ath Bilbao	Levante	Vallecano	Malaga
13	Almeria	Almeria	Zaragoza	Levante	Betis	Vallecano	Espanol	Espanol	Getafe	Granada
14	Zaragoza	Santander	Sociedad	Osasuna	Espanol	Betis	Valladolid	Osasuna	Espanol	Valladolid
15	Sp Gijon	Osasuna	Levante	La Coruna	Vallecano	Getafe	Granada	Ath Bilbao	Granada	Osasuna
16	Santander	Espanol	Getafe	Getafe	Zaragoza	Granada	Osasuna	Zaragoza	Elche	Elche
17	Malaga	Valladolid	Mallorca	Zaragoza	Granada	Santander	Celta	Granada	Almeria	Almeria
<b>Descenso</b>										
18	Tenerife	Zaragoza	La Coruna	Hercules	Villarreal	Villarreal	Mallorca	Celta	Osasuna	Getafe
19	Valladolid	Tenerife	Hercules	Almeria	Sp Gijon	Sp Gijon	La Coruna	Mallorca	Valladolid	Vallecano
20	Xerez	Xerez	Almeria	Malaga	Santander	Zaragoza	Zaragoza	La Coruna	Betis	Betis

Cuadro 7.2: Clasificaciones reales - puntuación ELO desde la temporada 2014/15 hasta la 2018/19

Posición	2014/15		2015/16		2016/17		2017/18		2018/19	
	Real	ELO								
<b>Campeón</b>										
1	Barcelona	Barcelona	Barcelona	Barcelona	Real Madrid	Barcelona	Barcelona	Barcelona	Barcelona	Barcelona
<b>UCL</b>										
2	Real Madrid	Real Madrid	Real Madrid	Real Madrid	Barcelona	Real Madrid	Ath Madrid	Ath Madrid	Ath Madrid	Ath Madrid
3	Ath Madrid	Real Madrid	Real Madrid	Real Madrid	Real Madrid					
4	Valencia	Valencia	Villarreal	Villarreal	Sevilla	Sevilla	Valencia	Valencia	Valencia	Getafe
<b>UEL</b>										
5	Sevilla	Villarreal	Ath Bilbao	Sevilla	Villarreal	Villarreal	Villarreal	Girona	Getafe	Sevilla
6	Villarreal	Sevilla	Celta	Ath Bilbao	Sociedad	Sociedad	Betis	Getafe	Sevilla	Valencia
<b>Media tabla</b>										
7	Ath Bilbao	Sociedad	Sevilla	Malaga	Ath Bilbao	Espanol	Sevilla	Celta	Espanol	Eibar
8	Celta	Malaga	Malaga	Eibar	Espanol	Ath Bilbao	Getafe	Villarreal	Ath Bilbao	Betis
9	Malaga	Espanol	Sociedad	Sociedad	Alaves	Eibar	Eibar	Eibar	Sociedad	Alaves
10	Espanol	Ath Bilbao	Betis	Celta	Eibar	Alaves	Girona	Sevilla	Betis	Sociedad
11	Vallecano	Celta	Las Palmas	Valencia	Malaga	Celta	Espanol	Sociedad	Alaves	Ath Bilbao
12	Sociedad	Vallecano	Valencia	La Coruna	Valencia	Las Palmas	Sociedad	Ath Bilbao	Eibar	Leganes
13	Elche	Getafe	Espanol	Betis	Celta	Valencia	Celta	Betis	Leganes	Villarreal
14	Levante	Eibar	Eibar	Las Palmas	Las Palmas	La Coruna	Alaves	Espanol	Villarreal	Girona
15	Getafe	La Coruna	La Coruna	Granada	Betis	Malaga	Levante	Leganes	Levante	Espanol
16	La Coruna	Almeria	Granada	Vallecano	La Coruna	Betis	Ath Bilbao	Alaves	Valladolid	Levante
17	Granada	Levante	Sp Gijon	Sp Gijon	Leganes	Leganes	Leganes	Levante	Celta	Celta
<b>Descenso</b>										
18	Eibar	Elche	Vallecano	Espanol	Sp Gijon	Granada	La Coruna	La Coruna	Girona	Huesca
19	Almeria	Cordoba	Getafe	Levante	Osasuna	Sp Gijon	Las Palmas	Malaga	Huesca	Vallecano
20	Cordoba	Granada	Levante	Getafe	Granada	Osasuna	Malaga	Las Palmas	Vallecano	Valladolid

## 7.1.2. Análisis de resultados

En la siguiente tabla de datos se muestran los distintos niveles de acierto que se han presentado durante las últimas 10 temporadas respecto a la base de datos de las clasificaciones reales. Se han presentado los distintos porcentajes de confianza en base a cada temporada y por cada grupo:

Cuadro 7.3: Porcentaje de acierto del sistema de puntuación ELO

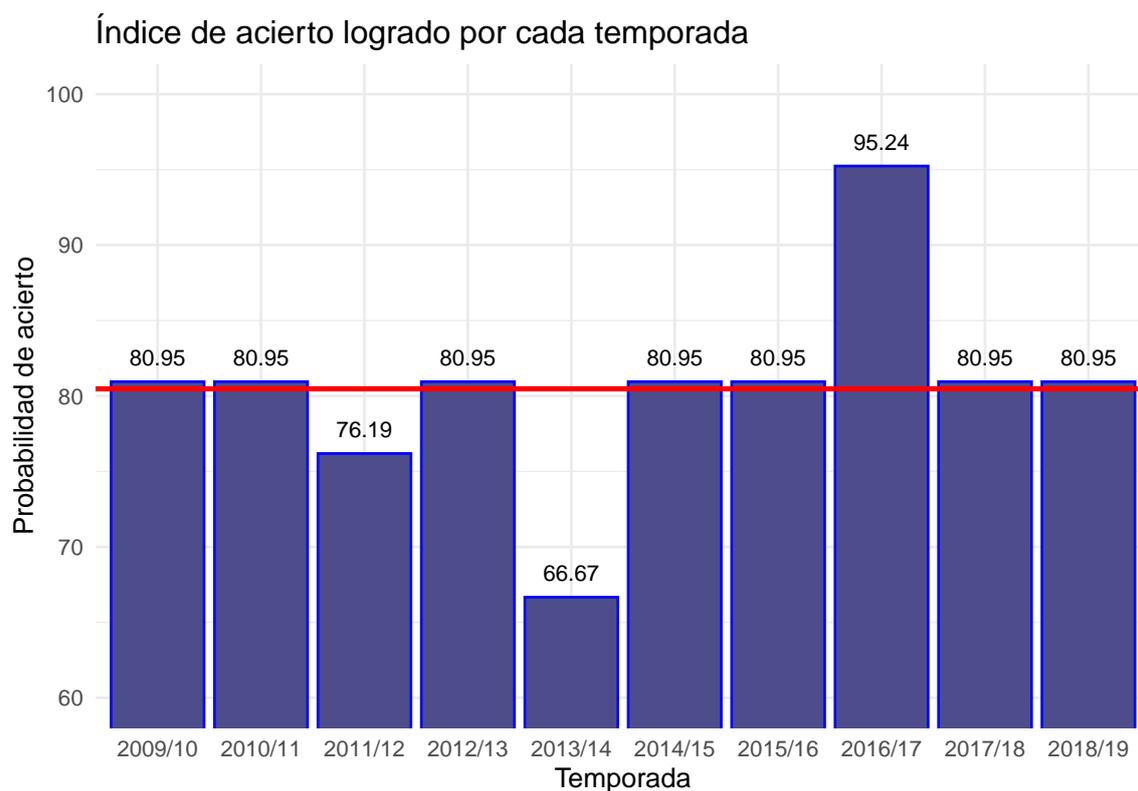
Grupo	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19	Total
Campeón	100.00	100.00	100.00	100.00	0.00	100.00	100.00	0.00	100.00	100.00	80.00
UCL	100.00	100.00	75.00	100.00	100.00	100.00	100.00	100.00	100.00	75.00	95.00
UEL	50.00	50.00	50.00	50.00	50.00	100.00	50.00	100.00	0.00	50.00	55.00
Media tabla	81.82	81.82	81.82	81.82	72.73	81.82	81.82	100.00	81.82	90.91	83.64
Descenso	66.67	66.67	66.67	66.67	33.33	33.33	66.67	100.00	100.00	66.67	66.67
Total	80.95	80.95	76.19	80.95	66.67	80.95	80.95	95.24	80.95	80.95	0.00

Indicar que la ponderación aplicada a la columna *Total* y a la fila *Total* han sido de la siguiente manera:

- Columna *Total*: media aritmética entre todos los datos de cada fila correspondiente.
- Fila *Total*: media ponderada de forma que se le asigna a cada grupo un peso específico en función al número de equipos que lo compongan. La tabla clasificatoria presenta 20 posiciones para los 20 equipos que disputan una temporada, pero dos grupos en concreto (*Campeón* y *UCL*) presentan el único solapamiento existente, donde coinciden el *Campeón* con el primer clasificado del grupo *UCL*, por lo que ponderamos a cada grupo sobre:

$$\text{Ponderación grupo} = \text{n}^{\circ} \text{ de equipos que conforman el grupo} / 21$$

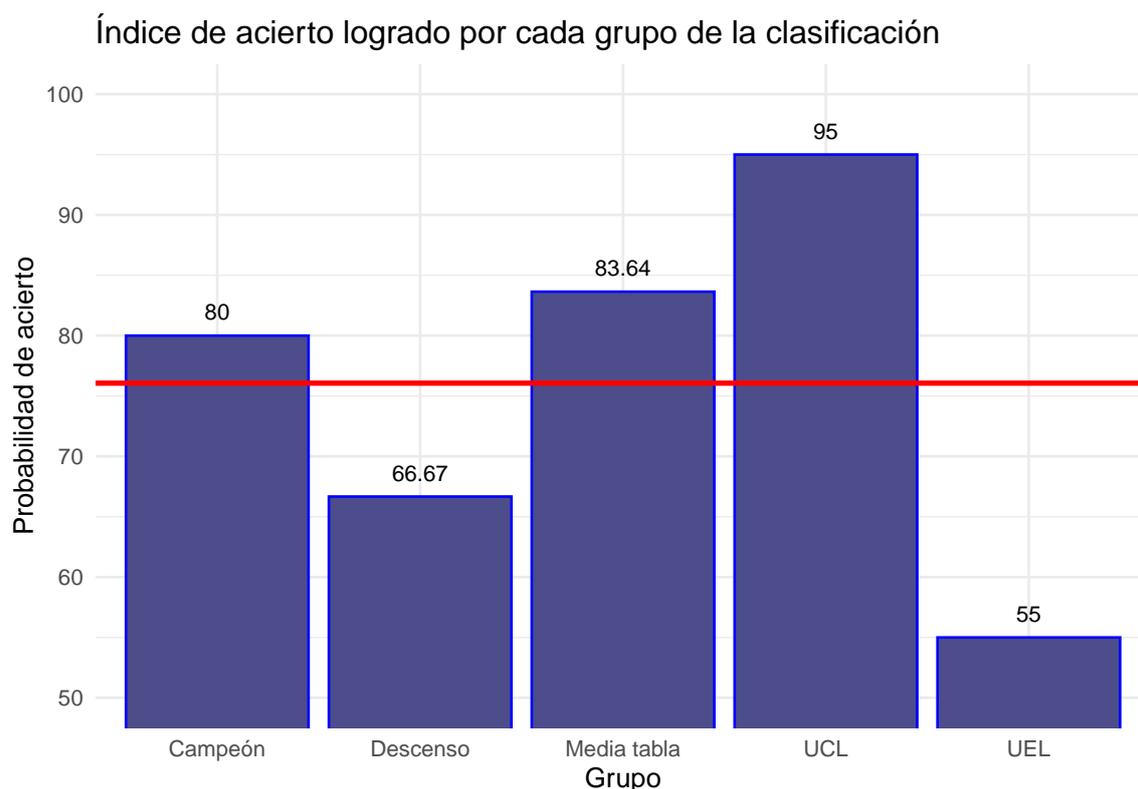
### 7.1.2.1. Porcentaje de acierto según temporadas



En esta ilustración damos figura a la representación gráfica del índice de acierto que tuvo el sistema de puntuación ELO en base a las verdaderas clasificaciones reales que se obtuvieron, manteniendo un muy buen nivel. La gran mayoría de temporadas se alcanza el acierto medio de un 80%, aunque hay algunos casos a resaltar como:

- Temporada 2013/2014, temporada en la que se presenta el peor porcentaje de acierto de todos. Tiene su cierta lógica pues esa fue la temporada en la que el Atlético de Madrid quedó campeón de LaLiga estando muy reñido con Real Madrid y Barcelona. Además, hubieron hasta 9 equipos en la lucha por no descender, lo que implica menos acierto de los equipos de *media tabla* y el grupo de *descenso*.
- En su opuesto se encuentra la temporada 2016/2017, en la que salvo algún que otro puesto que pudo estar algo reñido, el resto de grupos se encontraban medianamente definidos por sus equipos a falta de pocas jornadas, por lo que presenta un muy alto índice de acierto.

### 7.1.2.2. Porcentaje de acierto según grupos



Esta representación gráfica nos muestra el porcentaje de fiabilidad de acierto del sistema de puntuación ELO por cada uno de los grupos que hemos definido en la tabla clasificatoria, teniendo un acierto medio de un 76 % aproximadamente. Destacar que:

- El grupo de nuestra tabla de clasificación que presenta una clara mayoría de índice de acierto entre los equipos que lo componen es de aquellos equipos que participarán en la próxima campaña en la UCL. Este grupo presenta un registro de fiabilidad del 95 %, lo cual tiene bastante lógica puesto que la gran mayoría de años estos equipos suelen presentar cierta ventaja respecto al resto.
- Por otro lado, el grupo del descenso puede parecer que tiene un bajo nivel de acierto, pero para nada, puesto que hay que descontextualizar el grupo sabiendo el número de equipos que lo componen. Para este caso son 3 los equipos que conforman dicho grupo, por lo que presenta un nivel de fiabilidad considerablemente bueno. En el grupo de los equipos correspondientes a la UEL nos encontramos con un porcentaje de acierto del 55 %, lo que viene a ser 1 de cada 2 equipos cada temporada. Este nivel es bastante lógico puesto que las 2 únicas plazas de la UEL suelen estar mucho más disputadas que la UCL en casi todas las temporadas, habiendo casi siempre de 4 a 5 candidatos a final de temporada.
- Para finalizar nos quedan los grupos del campeón y de media tabla, con un nivel de confianza del 80 % y 83 % respectivamente. Ambos porcentajes no pillan de sorpresa, puesto que el campeón de LaLiga rara vez se define hasta las últimas jornadas y por ello tiene un índice de acierto tan alto. Por último, en el grupo de la media tabla, quitando algún equipo que se encuentre luchando por alguna plaza por poder disputar competiciones europeas o por mantener la categoría, el resto se encuentran en la zona tranquila de la tabla, ayudando a presentar un alto nivel de fiabilidad.

## 7.2. Real - Modelo

Como segunda comparación a realizar encontramos el bloque de la base de datos de las clasificaciones reales junto a la base de datos creada para nuestro modelo de predicción.

Para este caso no tenemos tablas de clasificación finales por temporada, puesto que como bien hemos aclarado antes, nuestro modelo predice en forma de clasificar a un equipo a uno de los distintos grupos de clasificación. De este modo no tenemos predicciones de resultados ni, por tanto, de puntuación y ranking clasificatorio.

Lo que sí obtenemos es el grupo al que ha sido asignado cada equipo en la última jornada de cada temporada, pudiendo saber qué equipos conforman cada uno de los mencionados grupos clasificatorios, pero el orden no tiene porqué ser el mismo.

### 7.2.1. Análisis de resultados

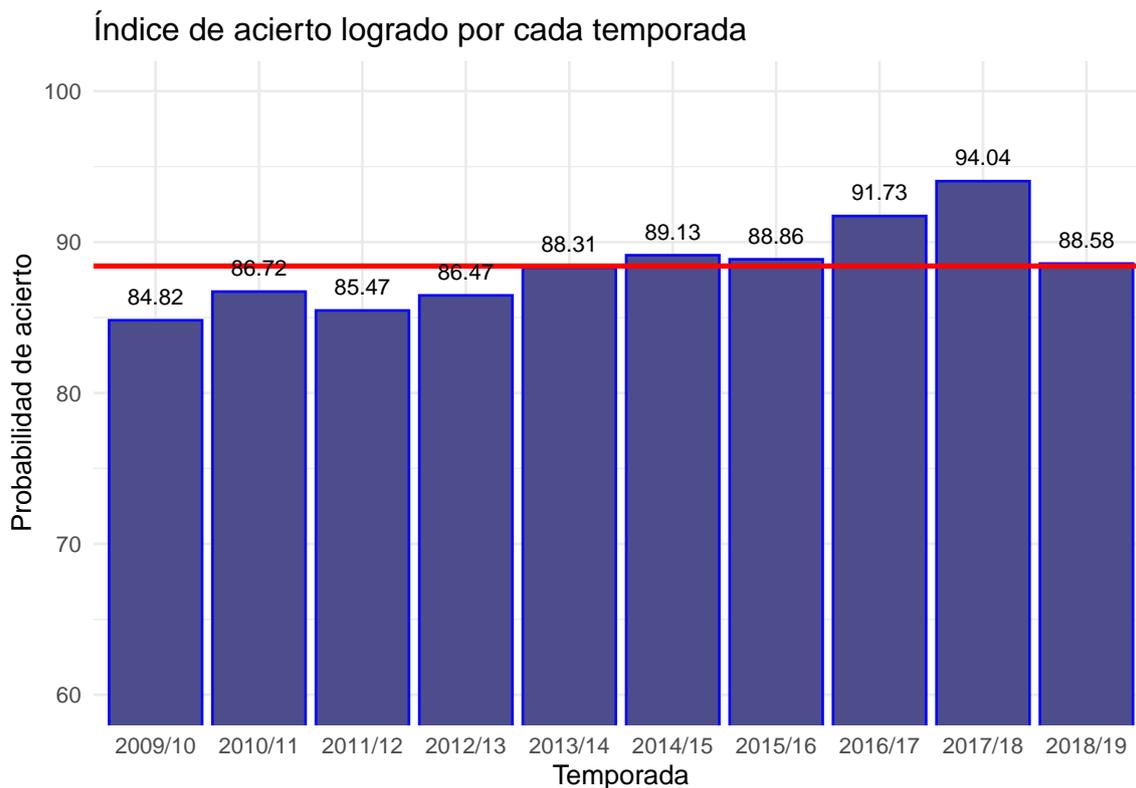
Los resultados que hemos obtenido provienen de la aplicación de nuestro modelo de predicción a cada una de las temporadas, extrayendo el índice de acierto que presenta por grupos y, a su vez, por temporadas. De esta manera podremos proceder a la medición de nuestro modelo con diversas representaciones gráficas y en forma de tablas. Aquí presentamos en el siguiente cuadro de datos el nivel de acierto mostrado por nuestro modelo de predicción en las últimas 10 temporadas disputadas:

Cuadro 7.4: Porcentaje de acierto de nuestro modelo de predicción

Grupo	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19	Total
Campeón	75.00	100.00	100.00	100.00	90.00	100.00	100.00	90.00	85.71	88.89	92.96
UCL	68.42	96.30	86.21	88.46	85.71	85.71	96.77	86.21	92.86	84.62	87.13
UEL	58.82	54.55	40.00	50.00	80.00	78.26	84.21	82.35	85.00	57.89	67.11
Media Tabla	95.79	91.84	88.18	89.72	93.20	92.59	88.24	94.39	96.19	93.40	92.35
Descenso	87.10	72.22	100.00	91.67	78.79	84.62	80.00	96.15	96.55	96.55	88.36
Total	84.82	86.72	85.47	86.47	88.31	89.13	88.86	91.73	94.04	88.58	0.00

Al igual que en el apartado anterior con la comparación de las clasificaciones finales reales con las clasificaciones obtenidas a través del sistema de puntuación ELO, hemos ponderado de igual manera la columna *Total* y la fila *Total* de este cuadro de datos, presentando los mismos criterios de cálculo.

### 7.2.1.1. Porcentaje de acierto según temporadas

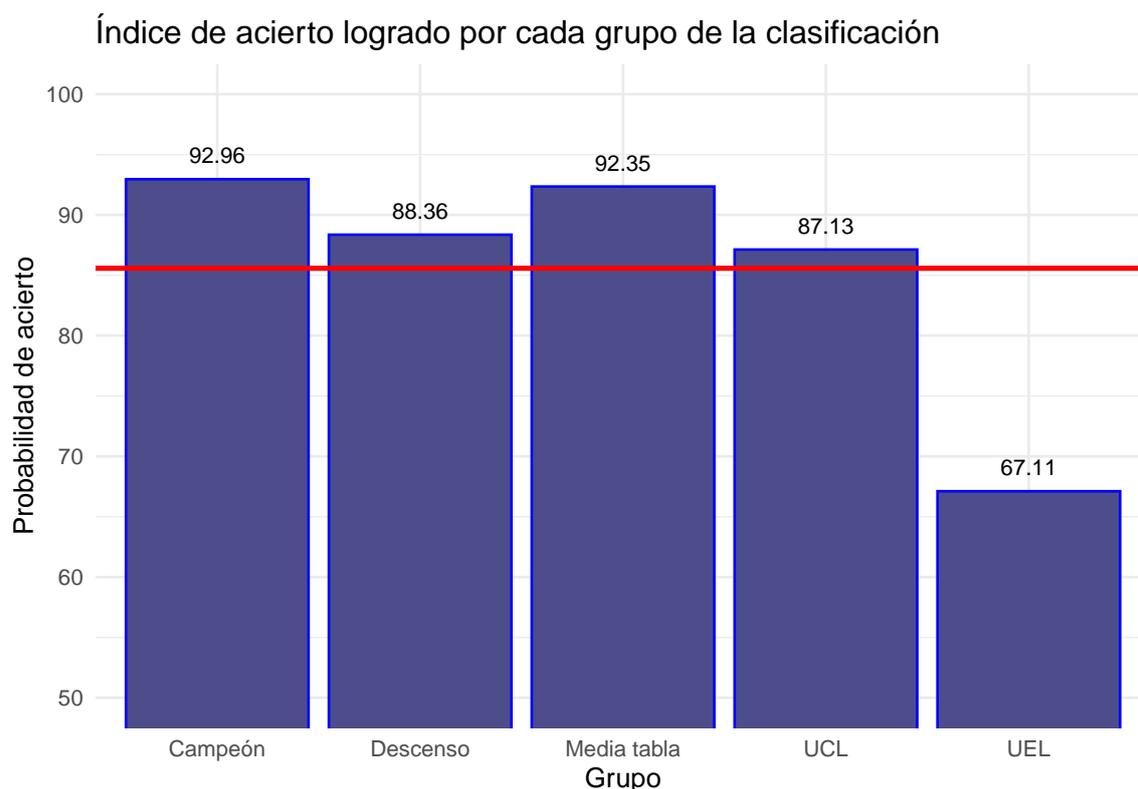


En este primer gráfico sobre nuestro modelo de predicción tenemos representado gráficamente el nivel de confianza mostrado en cada una de las temporadas, manteniendo siempre un nivel bastante alto. Algunos detalles o sutilezas a apreciar en el diagrama de barras son los siguientes:

- La media global de acierto clasificatorio está alrededor del 88 %, un nivel muy alto y con bastante fiabilidad.
- Con el paso de las temporadas se puede observar una ligera línea ascendente desde la temporada 2009/10 hasta la temporada 2017/18, donde casualmente han presentado los niveles más bajos (84.42 %) y más altos (94.04 %) respectivamente.

Como nota final podemos y creo especialmente que debemos hacer hincapié en el alto rendimiento que nos hemos encontrado, siendo un modelo de predicción de clasificación muy fiable y con mucha certidumbre.

### 7.2.1.2. Porcentaje de acierto según grupos



En esta representación gráfica se muestran los distintos porcentajes de acierto presentados para cada uno de los subgrupos de nuestra tabla clasificatoria, donde podemos resaltar algunos puntos:

- En todos los grupos se han obtenido unos niveles muy altos, ofreciéndonos un muy buen índice de certidumbre a la hora de clasificar a un equipo en un grupo u otro.
- Destacar el porcentaje logrado en el grupo de *UEL*, puesto que a priori al compararse junto al resto de índices de aciertos parece ser bajo. Sin embargo debemos notar que se queda a menos de un 3% para conseguir clasificar correctamente a 7 de cada 10 equipos que pertenecen al grupo en cuestión.
- Por último, indicar que se ha conseguido un promedio cercano o similar al del 86% de acierto; un muy alto listón sin duda.

## 7.3. ELO - Modelo

### 7.3.1. Tabla gráfica

A continuación presentamos una tabla donde aparecen las clasificaciones finales de la temporada 2019/2020 según cada uno de los dos modelos de predicción:

Cuadro 7.5: Clasificaciones obtenidas temporada 2019/2020

Posición	ELO		Modelo
	Equipo	Puntuación	Equipo
<b>Campeón</b>			
1	Real Madrid	1677.715	Real Madrid
<b>UCL</b>			
2	Barcelona	1671.002	Barcelona
3	Sevilla	1607.297	Sevilla
4	Getafe	1597.286	Ath Madrid
<b>UEL</b>			
5	Ath Madrid	1589.913	Getafe
6	Sociedad	1587.671	Sociedad
<b>Media tabla</b>			
7	Valencia	1528.137	Valencia
8	Ath Bilbao	1524.330	Ath Bilbao
9	Granada	1500.036	Granada
10	Villarreal	1491.962	Villarreal
11	Betis	1488.127	Betis
12	Osasuna	1481.759	Osasuna
13	Alaves	1451.393	Alaves
14	Levante	1429.919	Levante
15	Valladolid	1426.999	Valladolid
16	Celta	1416.458	Celta
17	Eibar	1415.744	Mallorca
<b>Descenso</b>			
18	Leganes	1387.176	Eibar
19	Mallorca	1386.866	Leganes
20	Espanol	1340.211	Espanol

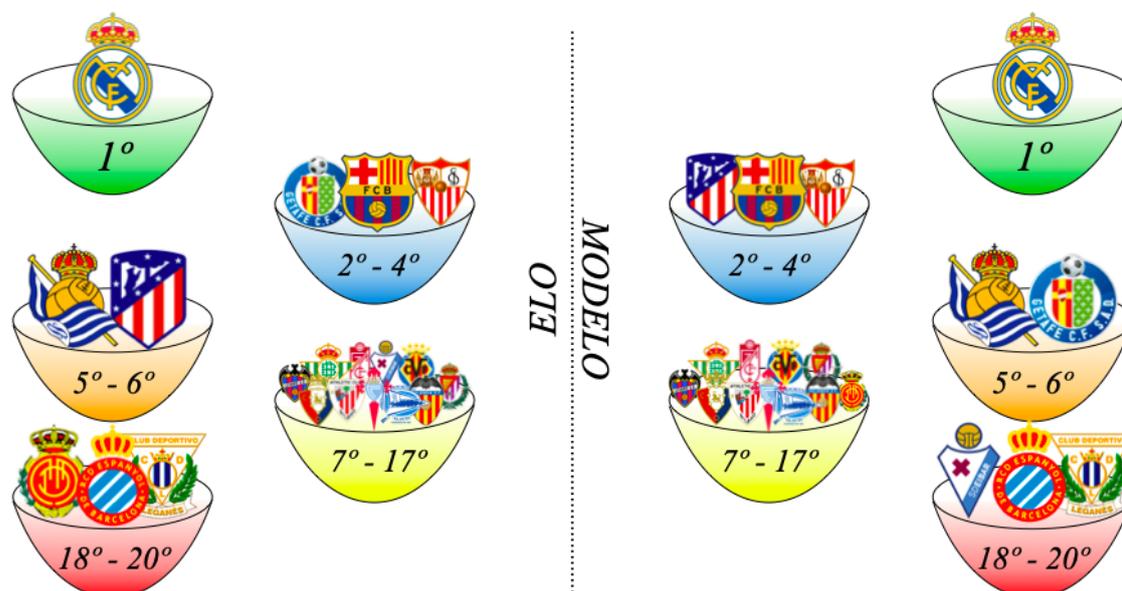
**NOTA:** Añadir que la clasificación final obtenida a través de nuestro modelo no se corresponde con su orden, es decir, cada uno de los equipos están correctamente formando el grupo al que pertenecen según han sido clasificados mediante nuestro modelo.

No obstante, la posición entre ellos puede variar respecto a la realidad, puesto que nuestro modelo se basa en la predicción de la clasificación de todos los equipos a cada uno de los distintos grupos y no de predicción de partidos en sí, por lo que no podemos obtener un ranking fijo como sí ocurre con el sistema de puntuación ELO.

Por último indicar que en la columna de *Puntuación* se puede ver la diferencia de puntuación entre los distintos equipos y dónde hay esos “saltos” de nivel entre ellos, puesto que a veces cuando hay uno o dos puestos de diferencia entre dos equipos puede haber una mucho mayor diferencia de puntuación ELO, indicándonos si un equipo es muy superior o no a otro.

### 7.3.2. Otra forma de representación gráfica

Además de la tabla anterior donde quedan reflejadas las clasificaciones obtenidas, aportaremos un esquema genérico con ambas clasificaciones finales de esta temporada 2019/2020 de una forma mucho más gráfica y visual.



En esta figura podemos apreciar varios detalles en cuanto a puestos clasificatorios de los equipos. A simple vista no hay muchas diferencias, como es el caso en que ambos procedimientos dan al Real Madrid como futuro campeón de LaLiga de la temporada 2019/2020. Sin embargo, hay otras sutilezas a marcar, como son:

- En el grupo de **UCL** hay una principal novedad; el Getafe y el Atlético de Madrid aparecen en dicho grupo en uno de los dos procedimientos. Si miramos la clasificación actual (hasta la jornada 27) podemos apreciar que ambos equipos se encuentran muy cercanos en la tabla clasificatoria, por lo que tiene sentido lógico que puedan diferir en la clasificación final de esta presente temporada.
- Por otro lado, estos dos mismos grupos se intercambian los papeles en el grupo de **UEL**, donde mediante el sistema de puntuación ELO el Atlético de Madrid es el que se queda sin plaza entre los 4 primeros clasificados. Por su contrario, nuestro modelo deja al Getafe como equipo perteneciente a la **UEL**.
- Para finalizar quisiera destacar el grupo del **Descenso**, donde ambas clasificaciones coinciden en dos de los tres equipos; Espanyol y Leganés. Sin embargo, tenemos una diferencia clave para el futuro de dos equipos como son el Mallorca y el Eibar, donde aparecen completando el grupo del **Descenso** en las clasificaciones logradas mediante el sistema de puntuación ELO y la obtenida por el modelo respectivamente.
- El resto de equipos lograrían el objetivo de mantener la categoría una temporada más en *LaLiga Santander* y así poder estar en la próxima campaña 2020/2021.

# Capítulo 8

## Conclusiones finales obtenidas

Como finalización de este Trabajo de Fin de Grado basado en la utilización de modelos estadísticos para la predicción de lo que queda de temporada 2019/2020 de LaLiga Santander podemos sacar unas conclusiones concluyentes en base a lo estudiado:

- Se ha logrado con bastante éxito en mi opinión el planteamiento, desarrollo y realización de cada una de las bases de datos distintas que hemos ido necesitando crear conforme hemos ido avanzando en el trabajo, por lo que se ha superado con cierta fiabilidad.
- Como objetivo principal teníamos conseguir una clasificación final de ambos métodos y poder así ver cuál era mejor o, por consiguiente, era más fiable. Según hemos visto podemos garantizar en cierta medida que ambos modelos han sido bastante resolutivos, cada uno con su planteamiento y funcionamiento, pero siendo muy fiables ambos.
- Podemos indicar que para la predicción de partidos como tal es mucho más fiable el modelo basado en la puntuación ELO, puesto que se basa en traducir cada encuentro en probabilidades de las posibles situaciones que nos podemos encontrar en el mismo. Sin menos mérito, podemos concluir que nuestro modelo de predicción es mucho más fiable en cuanto a la clasificación de un equipo a uno de los subgrupos de la tabla clasificatoria, puesto que es su única y especial función, además de los altos niveles presentados anteriormente.

Por último y no menos importante se ha llegado a una principal conclusión que engloba a todo este trabajo. La realización del mismo se ha trabajado en base a los precedentes de la competición (todos los encuentros de LaLiga de otras temporadas), y nos sería de mucha ayuda en caso de haber continuado la competición en su transcurso natural, como siempre se ha ido realizando cada campaña.

Debido a la delicada situación del COVID-19 esta experiencia es totalmente nueva para toda la gente que rodea al fútbol, desde los aficionados hasta los futbolistas. Ha sido un parón de unos tres meses en los que evidentemente cambian las tendencias de los equipos, jugadores, lesiones o fechas del calendario, puesto que muchos de los futbolistas hacen referencia a lo que queda por disputarse de competición como una “mini-liga” donde todos los equipos comienzan desde cero.

# Apéndice A

## Apéndice: Metodologías y herramientas

Para concluir este trabajo, se añadirá qué programas o aplicaciones han sido utilizados para llevar a cabo todo el proceso:

- **Draw.io:**

Esta aplicación puede utilizarse de forma online sin problemas desde tu cuenta de Google Drive, donde podrás llevar a cabo todo tipo de esquemas, diagramas UML o diagramas de entidad - relación entre otros. En mi caso lo he utilizado para el esquema realizado en el punto 6.3, donde se muestra gracias a Draw.io el esquema genérico de cada una de las variables del modelo de predicción que hemos creado.

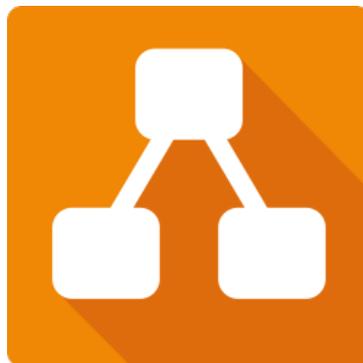


Figura A.1: Creación de esquemas o gráficos de entidad - relación entre otras: disponible en <https://www.draw.io/>

---

- **Rstudio:**

El software libre de RStudio es el gran artífice del trabajo, donde todo el resto del mismo ha sido creado a partir de él. Lo hemos utilizado para redactar todo el contenido y la creación de todas las tablas de datos, además de todos los gráficos e imágenes insertadas.

Si no fuese poco, recordar que la propia plantilla del trabajo está realizada exclusivamente por este software, ayudando a no necesitar ningún otro programa más y sin tener que gastar tiempo de trabajo en buscar programas o aplicaciones que te aporten cosas que RStudio ya te ofrece. Por último añadir que también tiene la opción de escribir en lenguaje LaTeX, ofreciendo todas las herramientas necesarias, además de contar con la generación automática de un PDF de altísima calidad.



Figura A.2: Software estadístico RStudio <https://rstudio.com>

# Bibliografía

- [1] Alejandro-Poza, E.A.P. 2010. *Wimbledon, el partido mas largo de la historia del tenis*. Disponible en <https://www.lawebdeltenis.net/wimbledon-el-partido-mas-largo-de-la-historia-del-tenis/>.
- [2] Álex-Merino, Á.M. 2020. *Ibai pide a william hill que quite las apuestas sobre su torneo, laliga challenge - movistar eSports*. Disponible en [https://esports.as.com/bonus/influencers/ibai/apuestas-aprovecha-torneo-benefico-Ibai\\_0\\_1338466145.html](https://esports.as.com/bonus/influencers/ibai/apuestas-aprovecha-torneo-benefico-Ibai_0_1338466145.html).
- [3] Calderón-Pérez, S.C. 2017. Disponible en [https://e-archivo.uc3m.es/bitstream/handle/10016/27480/TFG\\_Sergio\\_Calderon\\_Perez-Lozao\\_2017.pdf?sequence=1&isAllowed=y](https://e-archivo.uc3m.es/bitstream/handle/10016/27480/TFG_Sergio_Calderon_Perez-Lozao_2017.pdf?sequence=1&isAllowed=y).
- [4] Caro-Torres, Á.C. 2020. *Modelos estadísticos para la predicción de laliga 2019/20*.
- [5] casasdeapuestas.com *Los orígenes de las apuestas deportivas - casasdeapuestas.com*. Disponible en <https://www.casasdeapuestas.com/los-origenes-de-las-apuestas-deportivas/>.
- [6] Corporativa, R. 2019. *NotaPrensa*. Disponible en <http://www.funcas.es/prensa/NotasPrensa.aspx?file=399>.
- [7] C. Sergio-de la Cruz, S. de la 2020. *Los 38 acuerdos del futbol con casas de apuestas que 'salva' el real decreto de alberto garzon - elE*. Disponible en <https://www.economista.es/deporte-negocio/noticias/10370651/02/20/Los-38-acuerdos-con-casas-de-apuestas-del-futbol-que-salva-el-real-decreto-de-Alberto-Garzon.html>.
- [8] Daniel-Muñoz, D.M. 2017. *Reglas del ajedrez/ ¿como jugar al ajedrez?* Disponible en [https://thezugzwangblog.com/reglas-del-ajedrez/#En\\_que\\_casos\\_es\\_tablas\\_una\\_partida\\_de\\_ajedrez](https://thezugzwangblog.com/reglas-del-ajedrez/#En_que_casos_es_tablas_una_partida_de_ajedrez).
- [9] David-Salinas, D.S. 2020. *La liga española se paro durante la guerra civil*. Disponible en <https://www.sport.es/es/noticias/barca/liga-espanola-solo-paro-durante-guerra-civil-7890983>.
- [10] EFE 2020. *Estas son las profesiones mas demandadas en espana y en las que sobra empleo / el comercio*. Disponible en <https://www.elcomercio.es/economia/trabajo/profesiones-demandadas-espana-20200218024202-nt.html>.
- [11] ELDIAdigital.es 2020. *El drama de las casas de apuestas por el coronavirus / eldiadigital.es periodico de castilla-la mancha*. Disponible en <https://eldiadigital.es/art/324257/el-drama-de-las-casas-de-apuestas-por-el-coronavirus>.
- [12] elgurudelbasket.com *El partido mas largo de la historia de la nba el guru del basket*. Disponible en <https://elgurudelbasket.com/partido-mas-largo-historia-nba/>.

- [13] Emerging Technology From The Arxiv, traducido por M.C. 2017. *Unos científicos han descubierto una estrategia para ganar a las casas de apuestas 'online' | mit te*. Disponible en <https://www.technologyreview.es/s/9682/unos-cientificos-han-descubierto-una-estrategia-para-ganar-las-casas-de-apuestas-online>.
- [14] Ethan-Heinzen, E.H. 2020. *The "elo" package*. Disponible en <https://cran.r-project.org/web/packages/elo/vignettes/elo.html>.
- [15] FIFA.com 2017. *Clasificación mundial fifa/coca cola - clasificación completa - fifa.com*. Disponible en <https://es.fifa.com/fifa-world-ranking/ranking-table/men/rank/id11874/>.
- [16] FIFA.com 2017. *Clasificación mundial fifa/coca cola - noticias - suiza sube como la espuma - fifa.com*. Disponible en <https://es.fifa.com/fifa-world-ranking/news/suiza-subecomo-la-espuma-2902149>.
- [17] football-data.co.uk 2020. *Spain football results and betting odds*. Disponible en <https://www.football-data.co.uk/spainm.php>.
- [18] Hug, A.(. *Fifa-world-ranking-technical-explanation-revision*. Disponible en <https://resources.fifa.com/image/upload/fifa-world-ranking-technical-explanation-revision.pdf?cloudid=o1x9opdeidahofz8zbbp>.
- [19] ine *Introducción*. Disponible en [https://www.ine.es/explica/docs/historia\\_estadistica.pdf](https://www.ine.es/explica/docs/historia_estadistica.pdf).
- [20] Javier-Jiménez, J.J. 2018. *Científico de datos/ así es y así se forma uno en esta profesión cada vez más demanda*. Disponible en <https://www.xataka.com/otros/de-profesion-cientifico-de-datos>.
- [21] job.net 2020. *Ranking casas de apuestas | jobbet*. Disponible en <https://job.bet/ranking-casas-de-apuestas/>.
- [22] Jorge-Bautista, J.B. 2020. *LaLiga suspende la primera y la segunda división a causa del coronavirus*. Disponible en [https://www.lasexta.com/noticias/deportes/futbol/laliga-suspende-primera-segunda-division-causa-coronavirus\\_202003125e6a18227795c70001992bc9.html](https://www.lasexta.com/noticias/deportes/futbol/laliga-suspende-primera-segunda-division-causa-coronavirus_202003125e6a18227795c70001992bc9.html).
- [23] Jorge-Villar, J.V. 2020. *El gobierno decreta el estado de alarma*. Disponible en [https://www.lamoncloa.gob.es/consejodeministros/resumenes/Paginas/2020/14032020\\_alarma.aspx](https://www.lamoncloa.gob.es/consejodeministros/resumenes/Paginas/2020/14032020_alarma.aspx).
- [24] Luque-Calvo, P.L. 2017. *Escribir un trabajo fin de estudios con r markdown*. Disponible en <http://destio.us.es/calvo>.
- [25] none 2019. *Tinder detalla su algoritmo/ no empareja solo por el físico y ya no usa el famoso «elo»*. Disponible en [https://www.abc.es/familia/parejas/abci-tinder-detalla-algoritmo-no-empareja-solo-fisico-y-no-famoso-201903171243\\_noticia.html](https://www.abc.es/familia/parejas/abci-tinder-detalla-algoritmo-no-empareja-solo-fisico-y-no-famoso-201903171243_noticia.html).
- [26] Raúl-Rioja, R.R. 2019. *Las casas de apuestas/ un negocio millonario para los clubes de la liga*. Disponible en <https://www.20minutos.es/deportes/noticia/casas-apuestas-negocio-millonario-clubes-laliga-3661929/0/>.
- [27] Roberto-Ramajo, R.R. 2018. *La real sociedad dice 'no' a las casas de apuestas - as.com*. Disponible en [https://as.com/futbol/2018/12/21/primera/1545406991\\_087693.html](https://as.com/futbol/2018/12/21/primera/1545406991_087693.html).

[28] Sánchez-Casademont, R.S. 2020. *Ibai llanos traslada la liga a fifa por la cuarentena*. Disponible en <https://www.esquire.com/es/actualidad/a31735183/liga-fifa-ibai-llanos-courtois-carvajal/>.

[29] Sergio-Álvarez, S.Á. 2019. *¿Para que sirven los chalecos de los jugadores de futbol? - besoccer*. Disponible en <https://es.besoccer.com/noticia/para-que-sirven-los-chalecos-que-llevan-los-jugadores-en-los-partidos-16843> .