UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas /UPV

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas

Análisis del desequilibrio competitivo en las cinco grandes ligas europeas masculinas.

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Bosch Cervera, Alejandro

Tutor/a: Debón Aucejo, Ana María

CURSO ACADÉMICO: 2023/2024

Resumen

El fútbol, reconocido como el deporte con mayor seguimiento a nivel global, encuentra en las principales ligas masculinas de Europa —específicamente en España, Inglaterra, Alemania, Italia y Francia— un escenario de alta competitividad y lucratividad. No obstante, la disparidad existente entre los clubes que las integran repercute negativamente en el equilibrio competitivo, disminuyendo así el interés en dichos torneos.

Durante la elaboración de este proyecto se incidirá en el concepto de Balance Competitivo; esta medida expresa que, a mayor competitividad en una liga, mayor seguimiento e interés junto a unos posibles mayores ingresos generados a largo plazo. Se usará este concepto junto a un grupo de índices estadísticos elaborados, adaptados al sector del deporte para poder medir el desequilibrio competitivo existente en las temporadas 1999/2000 hasta la 2022/2023.

Para llevar a cabo este análisis se usará R Studio (RStudio Team, 2023), el cual mediante el paquete WorldFootballR (Zivkovic, 2024) permitirá obtener los datos comentados anteriormente por medio de la web Fbref (https://fbref.com/es/) mediante web scrapping.

Una vez formada la base de datos y seguido de la confección de los índices estadísticos, junto a otras variables, se ajustarán modelos de *Machine Learning* que permitirá conocer qué variables son las causas que determinan que en una liga exista mayor desequilibrio competitivo y por tanto una posible disminución de los ingresos a largo plazo.

Palabras clave: Fútbol; Ligas europeas; Desigualdad competitiva; WorldfootballR; Base de datos; Análisis estadístico; Modelos de machine learning; R Studio; Rentabilidad.

Abstract

Football, recognized as the sport with the largest global following, finds in the main men's leagues of Europe —specifically in Spain, England, Germany, Italy, and France— a scenario of high competitiveness and profitability. However, the existing disparity among the clubs comprising them negatively affects the competitive balance, thus diminishing interest in these tournaments.

Throughout the development of this project, emphasis will be placed on the concept of Competitive Balance. This measure expresses that the higher the competitiveness in a league, the greater the following and interest, along with potentially higher long-term generated revenues. This concept will be used alongside a set of elaborated statistical indices adapted to the sports sector to measure the existing competitive imbalance in the seasons from 1999/2000 to 2022/2023.

To carry out this analysis, R Studio (RStudio Team, 2023) will be used, which through the WorldFootballR package (Zivkovic, 2024) will allow obtaining the aforementioned data through web scraping from the Fbref website (https://fbref.com/es/)

Once the database is formed and statistical indices are constructed, along with other constant variables, a Machine Learning model will be implemented to determine which variables are the ones determining greater competitive imbalance in a league and therefore a potential decrease in long-term revenues.

Keywords: Soccer; European leagues; Competitive inequality; WorldfootballR; Database; Statistic analysis; Machine learning models; R Studio; Profitability.

Índice de contenidos

Índice de figuras

Índice de tablas

1. Introducción

El fútbol hoy en día es el deporte más seguido y popular del mundo, las ligas más interesantes y que mayores ingresos generan son las europeas, en concreto cinco (Inglaterra, España, Italia, Alemania y Francia).

Este deporte, es el más seguido y visto mundialmente, entre este destaca el fútbol europeo, en concreto las ligas comentadas junto a sus respectivos clubes, son las que mayores ingresos obtienen por derechos televisivos tanto a nivel nacional como internacional. (Fernández, 2019)

Es por lo que, por una parte, se da por hecho que el fútbol es un fenómeno social, pero también un negocio multimillonario en el cual los ingresos generados por los clubes y ligas, viene generado por contratos televisivos, patrocinios, merchandaising etc.

Estas dinámicas de mercado, junto a influencias de agentes económicos como los propietarios de clubes, patrocinadores y ligas influyen de manera directa en la competitividad de los equipos.

En Europa existen diversas ligas, pero solo son cinco las más grandes, esto es por su competitividad. La competitividad que existe en las ligas está relacionada con lo rentables que son. El deporte y el negocio no pueden separarse entre ellos hoy en día, sobre todo en el mundo profesional, el negocio muchas veces, aunque no quiera trata de obtener el máximo beneficio a costa de eliminar la competición, pero los clubs y ligas necesitan de competencia entre ellos para generar una rivalidad en el campo de calidad similar y brindar lo que quieren los aficionados, entretenimiento. (Stephen Dobson, 2011).

En este contexto, los organismos deportivos que supervisan las ligas juegan un papel fundamental en mantener una competencia equilibrada. Vrooman (2015) introduce el concepto del partido perfecto, donde los jugadores de diferentes equipos exhiben igual calidad y talento. Sin embargo, el desafío económico radica en el hecho de que las ligas de fútbol a menudo funcionan como cárteles naturales anticompetitivos. Estas ligas presencian equipos compitiendo con un poder económico asimétrico, creando una dinámica donde lograr el equilibrio ideal se convierte en un desafío sustancial.

1.1 Motivación

El análisis de la competencia en las diferentes ligas puede medirse con el Balance Competitivo, el cual se desarrollará más adelante.

Durante este Trabajo Fin de Grado (TFG), se elaborarán diferentes índices que analizarán el balance competitivo en las cinco grandes ligas europeas para de esa manera poder concluir gracias a un modelo de *Machine Learning* junto a los índices y otras variables que interactúan con ellos, cual es la más desbalanceada y por ende posiblemente menos rentable a largo plazo para el sector futbolístico.

1.2 Objetivos

El objetivo principal de este proyecto es conocer la liga más desequilibrada y por ende a largo plazo posiblemente menos rentable entre las cinco grandes ligas europeas.

Para lograr este objetivo será necesario:

- Obtener los datos de las cinco grandes ligas europeas, para las temporadas 1999/2000 hasta la 2022/2023 a partir del paquete WorldFootballR (Zivkovic, 2024), mediante la web Fbref (https://fbref.com/es/) y un código que se programará para usar la técnica de web scraping
- Elaborar distintos índices que ayudarán a entender el grado de desequilibrio competitivo en cada liga y temporada.
- Obtener el valor de mercado para cada equipo de cada liga y temporada, mediante el uso de la web Transfermarkt (https://www.transfermarkt.com/) utilizando de nuevo el paquete WorldFootballR. Estos valores se resumirán convenientemente para cada liga y temporada en su media y desviación típica.
- Elaborar modelos de Machine Learning mediante los índices y otras variables calculadas como la media y desviación típica. Posteriormente, aplicar los métodos de regresión de los algoritmos Random Forest y Regresión Lineal

1.3 Estructura del TFG

El presente TFG se ha estructurado de la siguiente manera:

En el capítulo 2 se encuentra el marco contextual o reconocimiento de trabajos de otros autores.

Este está compuesto por cuatro subsecciones: 2.1 El Balance Competitivo, 2.2 Estudios anteriores sobre el Balance Competitivo, 2.3 Las cinco grandes ligas europeas masculinas y 2.4 Recapitulación. En la primera se comentan las diferentes definiciones del Balance Competitivo, su importancia en este TFG y su contexto y como medirlo según distintos autores, entre otros aspectos. En segundo lugar, se comentan los estudios anteriores sobre el Balance Competitivo, en este caso en un contexto general. En tercer lugar, se desarrollará la evolución que han sufrido estas cinco grandes ligas desde su creación hasta día de hoy, además también se hablará sobre ciertos hitos que han afectado al valor de mercado de los equipos de estas ligas.

En el capítulo 3 se encuentra explicada la metodología utilizada en este TFG.

Este está formado por tres subsecciones: 3.1 Muestra, 3.1.1 Origen Datos, 3.1.2 Variables 3.2 Métodos estadísticos/econométricos/Aprendizaje Automático

En la primera se explica detalladamente cómo se construyó la muestra, que a su vez se distribuye en dos subsecciones: 3.1.1 Origen de los datos y 3.1.2 Variables. El punto 3.1.1 habla sobre cómo se han obtenido los datos y el proceso para obtenerlos y de esa manera confeccionar la muestra. Por su parte, el 3.1.2 explica de manera más detallada las variables que existen en la muestra.

En la segunda se describen los procesos estadísticos/econométricos/aprendizaje automático utilizados durante la elaboración de este TFG. Dentro de este se encuentra la subsección 3.2.1, donde se explican los índices estadísticos utilizados para medir el Balance Competitivo junto a otros procesos estadísticos elaborados, seguido de la subsección 3.2.2, donde se explican los modelos de aprendizaje automático (Machine Learning) de manera completa, siendo estos los que servirán para poder dar con la liga más desequilibrada competitivamente.

En el capítulo 4 se encuentran los resultados del análisis de los índices y del modelo de *Machine Learning*. Por tanto, encontraremos dos subsecciones 4.1 Análisis índices Balance Competitivo y 4.2 Análisis modelo *Machine Learning*.

En la primera se comentarán los resultados de los índices, descritos en las subsecciones: 4.1.1 SHICB, 4.1.2 Ratio de Desviación estándar. A continuación, dos subsecciones más 4.1.3 Medidas de porcentaje de empate, 4.1.4 Medidas de la dominancia, 4.1.5 Análisis de la evolución de los índices y 4.1.6 Análisis de correlación de los índices. Aquí a través de gráficas, tablas y de un resumen se expone el análisis de cada índice.

En la segunda se comentarán con más detalle los resultados del modelo de *Machine Learning*. En estos se encuentran los puntos 4.2.1 Análisis del modelo de Regresión Lineal, donde se expondrán los resultados del modelo, de manera similar que el punto 4.2.2 Análisis del modelo de Random Forest.

En el capítulo 5 se expondrán las conclusiones. Aquí se hará una síntesis de lo realizado y se justifica el cumplimiento de los objetivos. También se comentan las limitaciones encontradas durante la realización de este TFG y las perspectivas a futuro.

Por último, en la Bibliografía se encuentran todas las fuentes de información utilizadas durante el desarrollo del presente trabajo, además de poder consultarse también han sido citadas en formato APA.

2. Marco contextual

En este capítulo se contextualiza el trabajo realizado. Primero, se abordará el tema del Balance Competitivo, destacando su importancia en el ámbito futbolístico. Luego, se revisarán las investigaciones previas sobre el Balance Competitivo, enfatizando también su aplicación en el fútbol. Finalmente, se analizarán las cinco grandes ligas de fútbol masculino europeas, resaltando su evolución competitiva y económica a lo largo de los años.

2.1. El Balance Competitivo

A continuación, se procederá a desarrollar la importancia del Balance Competitivo y la importancia de conocer y entender este concepto para la elaboración de este TFG.

El Balance Competitivo es un concepto que distintos académicos definen con términos diferentes. Su estudio es difícil, pues se pueden encontrar diferentes medidas junto a interpretaciones distintas, sumado a que cada liga puede tener diferentes reglas o condiciones respecto al resto. La mayoría de los autores sí coinciden en que cuanta mayor competitividad existe en una liga, mayor atractivo para los aficionados y unos posibles mayores ingresos a largo plazo.

Silva (2018), afirma que el Balance Competitivo entre equipos en una competición se define por los ingresos generados por la taquilla, operaciones del estadio, patrocinios y los derechos de retransmisión realizados por los clubes participantes de la liga. Este autor incide más en el equilibrio competitivo en términos de ganancias económicas para los equipos involucrados

El término de Balance Competitivo, como se ha dicho, no es fácil de definir ni de medir. En este caso en términos de "cercanía de competición", este término aparece relacionado por primera vez en un artículo seminal de Economía del Deporte de Rottenberg (1956), desde ese momento ha sido objeto de atención en literatura especializada. Consecuencia de este hecho por lo que existen diferentes definiciones y maneras de medirlo.

Seguido del estudio de Rottenberg de 1956, el siguiente análisis más relevante y que ha provocado mayor interés en el tema del Balance Competitivo, ha sido el artículo de Walter C. Neale (1964), donde manifiesta la peculiaridad de la industria del deporte profesional, este designará y se referirá a sus características y necesidades como una economía peculiar. Neale plantea su "economía peculiar" de la industria del deporte como un mercado excepcional, puesto que ninguna empresa o, en este caso, equipo, podría funcionar como un monopolio o excluyendo a sus competidores del mercado.

En este artículo mediante la paradoja de Louis y Schmelling, se aborda la importancia que tiene la competición deportiva de existencia de contendientes de un nivel similar que aseguren el interés del público. Es decir, en un campeonato deportivo, es fundamental la presencia de dos o más equipos. Un solo equipo no puede generar suficiente interés ni beneficio por sí solo. La existencia de múltiples equipos crea una competencia que aumenta la emoción y la incertidumbre sobre quién será el ganador. Esto resalta la importancia del Balance Competitivo para garantizar un nivel de competencia equilibrado y emocionante para los espectadores.

El autor Ramchandani et al. (2018) manifestó que, según la literatura de equilibrio competitivo del modelo de deportes de equipos profesionales norteamericanos, algunos factores que afectan a la competencia entre los que se encuentran el reparto de ingresos, sistemas de Draft, es decir que las ligas tengan la oportunidad de seleccionar a los equipos los nuevos talentos a la lista de sus jugadores. También explica la necesidad de aplicar límites salariales y ligas cerradas para mantener el equilibrio competitivo.

En el caso de Plumley et al., (2018) este desarrolla la existencia de una fuerte vinculación entre los conceptos de incertidumbre de resultado, equilibrio competitivo y maximización de beneficios con los deportes de equipo profesionales. Relacionado con esto existen diferentes evidencias sustanciales entre los deportes de equipo norteamericanos y europeos en cuanto a términos de organización y estructura de liga, siendo estos unos factores que desempeñaron un papel importante en la forma de la literatura moderna entorno a la economía del deporte. Mientras que El-Hodiri y Quirk (1971) justifican las intervenciones regulatorias, siendo necesarias para poder maximizar el bienestar y mejorar el equilibrio competitivo, los autores Pawlowski y Budzinski (2014) expresan la probabilidad de que los aficionados o consumidores prefieran cierto desequilibrio, tal como se explica en la teoría económica de las superestrellas y el estrellato. En esta, Rosen (1981) sugiere que en ciertas industrias, un número reducido de individuos altamente talentosos o "superestrellas", reciben la mayor parte de los ingresos o atención del público. Esto se debe a que la demanda de productos o servicios asociados con estas superestrellas es extremadamente elástica en relación con su calidad o fama. El estrellato en este caso se refiere a la condición de ser reconocido y admirado en un campo específico, mientras que las superestrellas son los que han alcanzado el nivel máximo de estrellato, con el resultado de disfrutar de unos beneficios económicos significativos.

En este contexto se encuentran dos posibles suposiciones de comportamientos entre los equipos profesionales: el modelo norteamericano el cual busca maximizar los beneficios, y y el modelo europeo que busca obtener el mayor número de victorias en la temporada.

Dada la proliferación de investigadores que tratan el Balance Competitivo, ha desencadenado la aparición de diferentes vertientes, Fort y Maxcy (2016) identifican la clasificación de la investigación teórica y empírica del equilibrio competitivo en términos de: (I) análisis del balance competitivo, centrándose esta en lo que ha sucedido con el equilibrio competitivo a lo largo del tiempo como resultado de los cambios en las prácticas empresariales de las ligas deportivas profesionales; (II) La incertidumbre de la hipótesis del resultado, refiriéndose al análisis del equilibrio competitivo que mide el efecto en los aficionados.

Dentro del análisis del balance competitivo puede distinguirse en dos aspectos: (a) nivel de concentración y (b) nivel de dominio (Ramchandani et al., 2018). En este caso el nivel de concentración mide el grado de cercanía entre los equipos en una liga, mientras que el nivel de dominio se centra en la medida en que los mismos equipos continúan ganando la liga varias temporadas.

2.2. Estudios anteriores sobre el Balance Competitivo

Existe una gran cantidad de estudios que han cubierto el Análisis del Balance Competitivo, sobre todo, estos se han centrado en las ligas de deporte norteamericanas. Esto se puede atribuir al artículo seminal de Economía del Deporte de Rottenberg en 1956. Buzzacchi et al. (2003) junto a otros autores compara ambos modelos de deportes profesionales, en este estudio se analizó el número de equipos que tuvieron los porcentajes de victorias más altas en una temporada regular de la MLB, NFL y NHL. También se analizó el número de equipos que ganaron la liga de fútbol en Bélgica, Italia y Inglaterra entre los años 1950 y 1999. En este estudio se encontró que las ligas abiertas eran menos balanceadas que las cerradas en general. (Szymanski et al., 2003)

No fue hasta hace unos pocos años que el número de estudios centrados en equipos profesionales en Europa han aumentado, con un alto interés en el fútbol, siendo la mayoría, como en este caso, centrados en "las cinco grandes ligas europeas" (La Liga, Premier League, Bundesliga, Serie A y Ligue 1).

Estudios anteriores sobre el Balance Competitivo muestran algunos hallazgos contradictorios, en este caso algunos autores manifiestan la existencia de un decrecimiento en el equilibrio competitivo en las cinco grandes ligas europeas, como es el caso de Groot y Goossens (2008), (2006). En cambio, algunos manifiestan que no existe ningún cambio en el Balance Competitivo a lo largo de las temporadas analizadas, como es el caso de Szymanski, Mitchie y Oughton (2001).

Existen también otros enfoques académicos sobre el Balance Competitivo como es el caso de Ramchandani et al. (2019), en este se analiza el equilibrio competitivo en las fases de grupo de la UEFA Champions League, donde se encontraron ciertos defectos en el sistema de clasificación proporcionado por la UEFA. Además se demostró evidencia estadística de que históricamente los grupos de fases de la Champions League han sufrido de desequilibrio competitivo.

Por otro lado, se ha estudiado la relación con diferentes variables económicas como es el caso del PIB de un país y variables deportivas como la asistencia a las ligas del país en cuestión a las puntaciones de equilibrio competitivo de los rankings de naciones de la UEFA. (Rocke, 2019).

2.3. Las cinco grandes ligas europeas masculinas

Los inicios del fútbol se remontan a la Edad Media, aunque con fuertes raíces folclóricas. Este deporte pasó a ser un juego codificado y disciplinado entre los hombres de la Gran

Bretaña del siglo XIX. Posteriormente, el fútbol se extendió rápidamente a muchas partes del mundo, incluyendo Europa, Sudamérica y otros lugares. (Walvin, 2014)

Respecto a la creación de las cinco grandes ligas mencionadas, estas se remontan a principios del siglo XX. En el caso de España, esta se funda en 1929, por la Real Federación de Fútbol. La liga española comenzó con diez equipos, con equipos como el Barcelona, Real Madrid o Athletic, se buscó con su fundación, organizar un campeonato nacional. Por otro lado, en Alemania se funda la Bundesliga en 1963, esta fue una respuesta a la decepcionante actuación de la selección nacional en el Mundial de 1962, como propuesta para centralizar el talento del fútbol alemán en una liga profesional única y propia. En el caso de Italia, es una de las primeras ligas profesionales de fútbol en Europa, esta liga se remonta a 1898 como un campeonato regional, no es hasta 1929-1930 que se convirtió en una liga profesional. También es el caso de Francia, pionera en la creación de ligas profesionales en Europa, la Ligue 1 se funda en 1932. Por último, en Inglaterra, el caso de la Premier League es el más singular, ya que fue fundada en 1992, debido a que los equipos que conformaban la Football League First Division decidieron romper lazos con la Football League para aprovechar un beneficioso acuerdo de derechos televisivos.

En lo que respecta al tema de la competitividad en estas ligas, estas han experimentado cambios a lo largo del tiempo, como bien se ha manifestado en anteriores puntos. Los autores Haugen y Knut (2008) explican que las cinco grandes ligas están sufriendo una disminución de la incertidumbre de resultado, es decir cada vez los resultados de los partidos son más predecibles lo que afecta tanto al equilibrio competitivo, interés de los aficionados y a un largo plazo, una posible disminución de ingresos. Dentro de los factores culpables de este decrecimiento se encuentran: la evolución de la Champions League, este formato presentado en 1992, los ganadores de la última edición, el equipo inglés Manchester City recibió 81.4 millones de euros, excluyendo coeficientes de la UEFA y derechos televisivos. Por ejemplo el Real Madrid ganador de la edición 21/22 recibió casi 120 millones de euros (Bankinter, 2024), en este caso se expone el hecho de que los mismos clubs están ocupando siempre los mejores puestos en el ranking UEFA, así como en las últimas fases de la Champions, lo que manifiesta una dominancia financiera significativa para un grupo limitado de clubs, esta dominancia se traduce en un resultado de desbalance competitivo, hecho que está creciendo considerablemente en Europa en los últimos años.

Por otro lado, en ámbito económico, se aborda la cuestión de los derechos televisivos en Europa, los cuales han experimentado un considerable crecimiento en los últimos años. En el caso de la Premier, esta ha acordado unos ingresos de 6.7 billones de libras para las temporadas 2025-2029 (El País , 2024). En este sentido, se observa que los derechos televisivos no se distribuyen de manera equitativa entre los clubes, sino que los más importantes reciben considerablemente más que los más pequeños, lo que conlleva a empobrecer a estos últimos y resulta en un desequilibrio competitivo.

LALIGA

TEMPORADA 2022-23

IMPORTES RESULTANTES DEL REPARTO SIN AJUSTES POR ACUERDO ASAMBLEARIO PLAN IMPULSO - CVC *

	Ingresos	Obligaciones
ATHLETIC CLUB	64,31	-5,47
FUTBOL CLUB BARCELONA	155,10	-13,18
R.C.D. ESPANYOL DE BARCELONA, S.A.D.	51,24	-4,36
REAL MADRID CLUB DE FUTBOL	155,79	-13,24
CLUB ATLETICO DE MADRID, S.A.D.	120,43	-10,24
SEVILLA FUTBOL CLUB, S.A.D.	83,31	-7,08
REAL BETIS BALOMPIE, S.A.D.	70,93	-6,03
REAL SOCIEDAD DE FUTBOL, S.A.D.	65,91	-5,60
CADIZ CLUB DE FUTBOL, S.A.D.	45,64	-3,88
REAL CLUB DEPORTIVO MALLORCA, S.A.D.	45,04	-3,83
VALENCIA CLUB DE FUTBOL, S.A.D.	67,65	-5,75
CLUB ATLETICO OSASUNA	49,65	-4,22
ELCHE CLUB DE FUTBOL, S.A.D.	45,30	-3,85
VILLARREAL CLUB DE FUTBOL, S.A.D.	63,35	-5,38
REAL CLUB CELTA DE VIGO, S.A.D.	51,17	-4,35
RAYO VALLECANO DE MADRID, S.A.D.	45,97	-3,91
REAL VALLADOLID CLUB DE FUTBOL, S.A.D.	46,49	-3,95
GIRONA FUTBOL CLUB, S.A.D.	46,85	-3,98
GETAFE CLUB DE FUTBOL, S.A.D.	53,38	-4,54
UNION DEPORTIVA ALMERIA, S.A.D.	45,09	-3,83
TOTAL:	1.372,59	-116,67
TOTAL LALIGA EA SPORTS + LALIGA HYPERMOTION:	1.525,10	-129,63

IMPORTES RESULTANTES DEL REPARTO TRAS AJUSTES POR ACUERDO ASAMBLEARIO PLAN IMPULSO - CVC**

	Ingresos	Obligaciones
ATHLETIC CLUB	66,56	-5,66
FUTBOL CLUB BARCELONA	160,53	-13,65
R.C.D. ESPANYOL DE BARCELONA, S.A.D.	50,64	-4,42
REAL MADRID CLUB DE FUTBOL	161,24	-13,71
CLUB ATLETICO DE MADRID, S.A.D.	119,03	-10,40
SEVILLA FUTBOL CLUB, S.A.D.	82,34	-7,19
REAL BETIS BALOMPIE, S.A.D.	70,10	-6,12
REAL SOCIEDAD DE FUTBOL, S.A.D.	65,14	-5,69
CADIZ CLUB DE FUTBOL, S.A.D.	45,11	-3,94
REAL CLUB DEPORTIVO MALLORCA, S.A.D.	44,51	-3,89
VALENCIA CLUB DE FUTBOL, S.A.D.	66,86	-5,84
CLUB ATLETICO OSASUNA	49,07	-4,29
ELCHE CLUB DE FUTBOL, S.A.D.	44,77	-3,91
VILLARREAL CLUB DE FUTBOL, S.A.D.	62,61	-5,47
REAL CLUB CELTA DE VIGO, S.A.D.	50,58	-4,42
RAYO VALLECANO DE MADRID, S.A.D.	45,44	-3,97
REAL VALLADOLID CLUB DE FUTBOL, S.A.D.	45,95	-4,01
GIRONA FUTBOL CLUB, S.A.D.	46,30	-4,04
GETAFE CLUB DE FUTBOL, S.A.D.	52,75	-4,61
UNION DEPORTIVA ALMERIA, S.A.D.	44,56	-3,89
TOTAL:	1.374,10	-119,12
TOTAL LALIGA EA SPORTS + LALIGA HYPERMOTION:	1.525,10	-132,30*

Datos en millones de euros.
*Resultado de liquidar los derechos audiovisuales conforme a los criterios del RD.
**Resultado de aplicar sobre la columna precedente los pagos y detracciones previstos en el acuerdo de la asamblea general de LALIGA de 10 de diciembre de 2021. Plan Impulso.

Figura 1. Reparto de los Derechos de TV LALIGA

Fuente: LALIGA (2024).

En la siguiente Figura 1, se puede observar cómo el reparto es desigual, como es el caso del Real Madrid, con mayores ingresos, frente a clubes como el Almería o Mallorca que reciben 4 veces menos derechos que el Real Madrid, fomentando un desequilibrio competitivo considerable. En el marco de este contexto, se vincula con el valor de mercado de los equipos. Como se ha explicado existe un desequilibrio a la hora de distribuir los derechos televisivos, lo que afecta entre otras causas al valor de mercado de los equipos, ya que estos derechos permiten que los partidos lleguen a cierta audiencia global, permitiendo una mayor visibilidad de equipos y jugadores, afectando a su valor de mercado. Por otro lado, esta disparidad fomenta las desigualdades de valores de mercado entre clubes, ya que los que mayores derechos reciben podrán comprar jugadores con mayor talento y por ende mayor valoración. El valor de mercado será un factor de estudio durante este TFG, más adelante se introducirá cómo se obtiene y cómo se integrará al modelo de *Machine Learning*.

Continuando con los factores que propician el desequilibrio, se encuentra el Fair Play Financiero impuesto por la UEFA. La idea detrás de esta medida en el fútbol europeo es dificultar que los clubes ricos gasten más para comprar talento. Esta normativa establece que los clubes no pueden gastar más de lo que ingresan. (Universidad Europea, 2023)

Por último, la introducción del sistema de puntuación 3-1-0, con el propósito de fomentar el juego ofensivo y por ende visualizar un fútbol más dinámico y entretenido, con muchos goles para el aficionado. Este sistema puede tener efectos adversos en el equilibrio competitivo, con una posible deriva a un juego más defensivo y aburrido para el aficionado.

2.4. Recapitulación

El Balance Competitivo es un concepto que ha sido abordado por varios académicos, aunque su definición y medición son desafiantes debido a la diversidad de enfoques y medidas utilizadas, así como a las diferentes condiciones en las ligas deportivas. A pesar de estas diferencias, la mayoría coincide en que una mayor competitividad en una liga atrae a más aficionados y potencialmente genera mayores ingresos a largo plazo.

Silva (2018) define el Balance Competitivo en términos de ingresos generados por los clubes participantes, destacando la importancia del equilibrio económico entre los equipos. Desde el seminal artículo de Rottenberg (1956), se ha prestado atención al concepto, con Neale (1964) enfocándose en la peculiaridad de la industria del deporte profesional y la necesidad de competencia equilibrada.

Autores como Ramchandani et al. (2018) y Plumley et al. (2018) exploran factores como el reparto de ingresos y la regulación salarial para mantener el equilibrio competitivo. Mientras que El-Hodiri y Quirk (1971) abogan por intervenciones regulatorias para mejorar el bienestar y la competencia.

Se discute la relación entre el modelo norteamericano, que busca maximizar beneficios y el europeo, que busca victorias. Los estudios sobre Balance Competitivo se clasifican en análisis del balance a lo largo del tiempo y la incertidumbre de resultados.

Se han realizado numerosos estudios sobre el tema, principalmente en ligas norteamericanas, aunque recientemente ha aumentado el interés en Europa, especialmente en el fútbol. Estudios anteriores muestran hallazgos contradictorios sobre el equilibrio competitivo en las ligas europeas.

Las cinco grandes ligas europeas masculinas (La Liga, Premier League, Bundesliga, Serie A y Ligue 1) tienen una historia que se remonta al siglo XX, cada una con su propia evolución. Sin embargo, estudios recientes sugieren una disminución en la incertidumbre de resultados, lo que afecta al equilibrio competitivo y potencialmente los ingresos a largo plazo.

Factores como la evolución de la Champions League, la distribución desigual de los derechos televisivos, *Fair Play* Financiero y el sistema de puntuación 3-1-0 han contribuido a un posible desequilibrio competitivo, con los clubes más grandes beneficiándose en detrimento de los más pequeños.

3. Metodología

En este capítulo se presenta y describe la metodología empleada en el TFG. La manipulación, preprocesamiento, modelado y análisis de datos fueron realizadas mediante el lenguaje de programación R (R Core Team, 2023) a través de la interfaz R-Studio (RStudio Team, 2023), a excepción de las variables económicas, las cuales fueron elaboradas mediante hojas de cálculo de Excel.

3.1. Muestra

La base de datos consta de las cinco grandes ligas europeas, en este caso, se han obtenido los datos para cada liga de cada país, para España, La Liga; para Inglaterra, la Premier League; para Alemania, la Bundesliga; para Italia, la Serie A; y para Francia, la Ligue 1. Dentro de cada dataframe de cada liga se encuentran 480 observaciones para LaLiga y Premier, 432 observaciones para la Bundesliga, 470 para la Serie A y 474 para la Ligue 1, todos con 16 variables. Existen dos bloques de variables, las relacionadas con el balance competitivo y las económicas, estas últimas corresponden a un dataframe con 24 observaciones y 11 variables. Durante el análisis se han utilizado los datos procedentes de las cinco grandes ligas para las temporadas 1999/2000 hasta la 2022/2023.

3.1.1. Origen de datos

Los datos correspondientes a las variables relacionadas con el balance competitivo se han obtenido a partir del paquete de R WorldFootballR (Zivkovic, 2024), mediante éste se ha podido junto a un código de web scraping, obtener los datos de la web Fbref (https://fbref.com/es/), mediante esta web se han extraído los datos para cada liga desde la temporada 1999/2000 hasta la 2022/2023. La información de los dataframes obtenidos por la web consiste en la clasificación para cada liga y temporada, junto a la posición y nombre del equipo correspondiente, puntos conseguidos, victorias, empates, derrotas entre otros, siendo estos los datos más relevantes para el análisis. Los datos de las temporadas mencionadas fueron descargados a fecha de 12 de febrero de 2024.

Se han creado unos índices que medirán el Balance Competitivo en la liga y temporada determinada a partir de los datos obtenidos por Fbref. Las fórmulas de cálculo de estos índices se detallan en la subsección 3.2.1.

En cambio, las variables económicas se han obtenido manualmente mediante la web Transfermarkt (https://www.transfermarkt.com/), todo para cada liga y su correspondiente temporada. Los datos de valor de mercado fueron descargados a fecha de 15 de marzo.

3.1.2. Variables

Las variables utilizadas durante el análisis son las siguientes:

Variables relacionadas con el balance competitivo:

- **Liga (liga):** variable categórica que identifica sobre qué liga están representadas las demás variables. Esta variable abarca: “ESP”, (España), “ENG” (Inglaterra), “ITA” (Italia), “GER” (Alemania) y “FRA” (Francia).

- **Temporada (temporada):** variable numérica discreta que identifica a qué temporada corresponde cada año y demás variables, en el periodo 2000 al 2023.

Índices que miden el Balance Competitivo:

- **Índice Herfindahl del Balance Competitivo (HICB):** El HICB es una versión normalizada del Índice Herfinfindahl-Hirschman (HHI) y otras medidas de dominancia como el número de diferentes equipos que ganan el título de liga o las diferentes posiciones en la clasificación de los equipos, para medir el balance competitivo. Corresponde a una variable numérica continua. Esta medida se atribuye a Mitchie (2004) y en el enfoque europeo mayormente a Ramchandani (2018)

- **Índice Herfindahl del Balance Competitivo Estandarizado (SHICB):** El SHICB corresponde a una versión estandarizada del HICB. Esta normalización se usa para poder medir de manera más equitativa en posteriores análisis a las diferentes ligas. Corresponde a una variable numérica continua. Esta medida se usa en el trabajo realizado por Mondal (2023)

- **Ratio Desviación Estándar (RSD):** El RSD corresponde a una variable numérica continua. Esta variable permite conocer la ratio de dispersión de victoria en una liga y temporada determinada. Esta medida se atribuye a Noll (1974). Corresponde a una variable numérica continua.

- **Porcentaje de empates (Draw_porcentaje):** Porcentaje de empates en una liga y temporada determinada. Corresponde a una variable numérica continua. Esta medida se atribuye a Kringstad (2018).

Variables económicas:

- **Valor Media del Valor de Mercado (Valor_Media):** Valor obtenido a partir de la media de valores de mercado de cada club en una liga y temporada determinada. Corresponde a una variable numérica continua.

- **Valor Desviación Típica del Valor de Mercado (Valor_DesvTip):** Valor obtenido a partir de la desviación típica del valor de mercado para cada club en una liga y temporada determinada. Corresponde a una variable numérica continua.

3.2. Métodos estadísticos / econométricos / aprendizaje automático

El análisis del balance competitivo se ha cuestionado a fondo en la industria de la gestión deportiva, dado que surgieron dudas sobre la precisión de la evidencia empírica debido a la gran cantidad de diferentes medidas para evaluar el equilibrio competitivo y su diferente evolución.

Se ha usado una extensa lista de variables como la desviación estándar de los porcentajes de victoria, el coeficiente de Gini, Curva de Lorenz o margen medio de victoria, la relación de equilibrio competitivo y diferentes medidas de concentración y dominio se han utilizado para analizar el equilibrio competitivo. Este TFG se ha centrado en realizar el análisis correspondiente con un enfoque europeo, como manifiesta Ramchandani (2018).

Dentro de este análisis se muestra la cantidad de equipos que existen en cada liga durante el periodo de tiempo que se realiza investigación.

Temporada	Alemania	España	Francia	Inglaterra	Italia
2000	18	20	18	20	18
2001	18	20	18	20	18
2002	18	20	18	20	18
2003	18	20	20	20	18
2004	18	20	20	20	18
2005	18	20	20	20	20
2006	18	20	20	20	20
2007	18	20	20	20	20
2008	18	20	20	20	20
2009	18	20	20	20	20
2010	18	20	20	20	20
2011	18	20	20	20	20
2012	18	20	20	20	20
2013	18	20	20	20	20
2014	18	20	20	20	20
2015	18	20	20	20	20
2016	18	20	20	20	20
2017	18	20	20	20	20
2018	18	20	20	20	20
2019	18	20	20	20	20
2020	18	20	20	20	20
2021	18	20	20	20	20
2022	18	20	20	20	20
2023	18	20	20	20	20

Tabla 1. Tamaño de las ligas por temporada

Fuente: Elaboración propia con los datos de Fbref y R.

Como se mencionó anteriormente, se ha realizado el análisis de las cinco grandes ligas con un enfoque europeo, por lo que se usarán en primer lugar ciertos índices estadísticos para evaluar el estado del balance competitivo en cada liga y temporada. Seguidamente

se construirá un modelo, el cual permitirá establecer relaciones entre el balance competitivo y varios factores que afectan a este. El ajuste del modelo se realizará mediante el uso de los métodos de ajuste de regresión lineal y el algoritmo Random Forest.

3.2.1. Métodos estadísticos

En primer lugar, se ha utilizado el Índice HICB. Como se explicó anteriormente, el HICB es el Índice de Herfindahl (HHI) aplicado al balance competitivo de una liga de fútbol. Una ventaja del HICB es que considera el número de equipos para calcular el Balance Competitivo, y puede emplearse tanto para comparar la misma liga en diferentes temporadas como para analizar distintas ligas con diferentes números de equipos.

$$\text{HICB} = \left(\frac{\text{HHI}}{\frac{1}{\text{N}}}\right) \times 100$$

Donde:

- *HHI* corresponde a la suma de los cuadrados de los puntos compartidos de cada club que disputa una liga en una temporada determinada
- *N* corresponde al número de equipos en esa liga y temporada concreta.

  Por ejemplo, si existen tres equipos en una liga, digamos A, B y C, con el equipo A ganando 4 puntos, B ganando 3 puntos y C ganando 1 punto el resultado del HICB calculado es:

$$HICB = \left\{\left[\left(\frac{4}{8}\right)^2 + \left(\frac{3}{8}\right)^2 + \left(\frac{1}{8}\right)^2\right] / \left(\frac{1}{3}\right)\right\} \times 100$$
$$= \left(\frac{0.40625}{\frac{1}{3}}\right) \times 100 = 121.87$$

Si el valor del $HICB$ es 100, es una liga perfectamente balanceada para cualquier tamaño. Cuanto más alto el HICB, menor Balance Competitivo.

Dado que el límite superior del HICB puede variar de acuerdo con la cantidad de equipos en una liga, se ha optado por calcular la versión estandarizada del HICB, llamada SHICB.

Se calcula como:

$$SHICB = \left(\frac{HICB}{\max H\,ICB(N)}\right) \times 100$$

Donde:

- Se tiene el $HICB$ para una liga y temporada determinada.

- Se tiene el $Max\ HICB\ (N)$, correspondiente al límite superior de la puntuación del HICB en la liga y temporada correspondiente, N corresponde al número de equipos.
- Para calcularlo, se realiza como: $HHI_{max} = \frac{2(2N-1)}{3N(N-1)}$
- Seguidamente se realiza: $\frac{HHI_{max}}{1/N}$

Un resultado de 100 para el $SHICB$ presenta una posición desbalanceada para esa liga. Cuanto menor es el SHICB, más balanceada o mayor balance competitivo.

Por otro lado, se ha calculado la ratio de dispersión de victoria en cada liga. Es una medida sesgada para calcular el Balance Competitivo, debido a la alta proporción de partidos empatados. Para mitigar este sesgo se sigue la forma tradicional de tratar los empates como medias victorias.

Se calcula como:

$$RSD = \frac{ASD}{ISD}$$

Donde:

- ASD corresponde a la desviación estándar del porcentaje de victorias al final de temporada
- ISD corresponde a la desviación estándar idealizada y se calcula como:

$$ISD = \frac{0.5}{\sqrt{m}}$$

- Donde m, corresponde al número de partidos jugados por cada equipo en la liga y temporada determinada.

Un resultado del RSD bajo indica un mayor equilibrio competitivo en la liga.

Por último, se ha calculado el Draw%, este índice es una medida que permite conocer el porcentaje de empates para una liga y temporada determinada.

Se calcula como:

$$Draw\% = \frac{\sum Empates}{(m \times N)}$$

Donde:

- $\sum Empates$, corresponde al sumatorio de de empates del equipo 1 al equipo N de una liga y temporada determinada.
- $(m \times N)$, corresponde a m que indica el número de partidos jugados para cada equipo en una liga y temporada determinada y N indica el número de equipos.

Un resultado de $Draw\%$ mayor implica que los equipos están muy igualados entre sí y por tanto existe una mejor presencia de equilibrio competitivo en una liga.

Se ha llevado a cabo un proceso estadístico con el propósito de analizar en profundidad los índices mencionados anteriormente. Este proceso implica la exploración de la evolución de estos índices, una etapa crucial para comprender las relaciones y tendencias presentes en los datos. Mediante esta exploración, se busca identificar posibles patrones y tendencias en los diversos índices estadísticos utilizados para evaluar el balance competitivo en las ligas de fútbol europeas. Este análisis de la evolución proporciona una visión más completa de cómo estos índices cambian a lo largo del tiempo, arrojando luz sobre la dinámica y la naturaleza de la competencia en el fútbol europeo. Este proceso estadístico de visualización de la evolución de los índices proporciona una base sólida para el análisis y la interpretación de la competitividad en las ligas de fútbol europeas, ayudando a identificar tendencias significativas.

Continuando con la visualización, también se ha elaborado un gráfico de dominancia en las cinco grandes ligas, donde en el eje X se observa el equipo con más campeonatos ganados, mientras que en el eje Y se muestran los diferentes equipos que han ganado el campeonato. De esta manera, se puede conocer el grado de dominio de los equipos en esa determinada liga a partir de las temporadas analizadas.

Finalmente, se llevará a cabo una correlación entre los índices, lo que permitirá identificar de manera resumida y clara cómo se relacionan entre sí, proporcionando una comprensión más profunda de los factores que influyen en la competitividad de las ligas.

3.2.2. Métodos de Aprendizaje Automático

Para complementar el análisis estadístico previamente mencionado, se construirá un modelo ajustado por dos métodos de Machine Learning: regresión lineal y Random Forest. Este enfoque permitirá no solo examinar el balance competitivo en las ligas europeas, sino también identificar y cuantificar el impacto de diversos factores, incluyendo variables económicas, en dicho balance.

Para obtener el dato del Valor de Mercado para cada temporada y liga se ha consultado la web de Transfermarkt, obteniendo así los datos para las temporadas 2005 hasta la 2023. No se ha podido obtener de ninguna otra manera los datos de las anteriores temporadas, por lo que se obviarán una vez se cree el modelo de *Machine Learning*. Para obtener la media, se ha dividido el valor de mercado de la liga y temporada determinada entre la cantidad de equipos que participan en ella. Para la desviación típica, se ha utilizado la función Desvest.M en Excel, para cada equipo de cada liga y temporada determinada.

En cuanto al Machine Learning, este trabaja en la creación de algoritmos y modelos que posibilitan a los ordenadores aprender de los datos y perfeccionar su funcionamiento a medida que adquieren experiencia, sin requerir una programación directa para cada tarea. En lugar de recibir instrucciones precisas, estos modelos emplean métodos estadísticos y computacionales para identificar patrones en los datos, lo que permite realizar predicciones o tomar decisiones fundamentadas en dichos patrones. Todo esto será posible con el uso de regresión lineal y la aplicación del algoritmo Random Forest

Para ello se usará el siguiente modelo:

$$\text{HICB o SHICB o RSD o Draw\%} = \beta_0 + \beta_1 \times O + \beta_2 \times P + \beta_3 \times Y + \beta_4 + \epsilon$$

Donde:

- O, es una variable económica explicativa que indica la media del valor de mercado para una temporada y liga determinada
- P, es una variable económica explicativa que indica la desviación típica del valor de mercado para una temporada y liga determinada
- Y, representa la tendencia del tiempo.
- El HICB, SHICB, RSD y DRAW%, son variables dependientes en el modelo de *Machine Learning*.

Una vez estén los datos de índices y variables se aplicará el método de regresión lineal a este modelo.

La regresión lineal es un método estadístico y de Machine Learning utilizado para modelar la relación entre una variable dependiente (denotada como y) y una o más variables explicativas (denotadas como $x_1, x_2,\ldots,x_n$)

Este método asume una relación lineal entre estas variables, expresada por una ecuación de la forma:

$$y=\beta_0+\beta_1 \cdot x_1+\beta_2 \cdot x_2+\ldots+\beta_n \cdot x_n+\varepsilon$$

Donde β_0 es la intersección y representa el valor de y cuando las variables valen cero, β_1, $\beta_2\ldots,\beta_n$ son los coeficientes que representan el cambio en y por cada cambio unitario en cada variable explicativa cuando el resto permanece constante, y ε es el término de error.

El objetivo es encontrar los valores de los coeficientes que minimicen la suma de los cuadrados de las diferencias entre los valores predichos y los valores reales de la variable dependiente en el conjunto de datos. Esto se logra mediante el método de mínimos cuadrados, que implica calcular las derivadas de la función de error con respecto a cada coeficiente, igualarlas a cero y resolver para obtener los valores óptimos de los coeficientes.

Los coeficientes $\beta_1, \beta_2,\ldots,\beta_n$ representan la magnitud y la dirección de la relación entre las variables explicativas y la variable dependiente. Un coeficiente positivo indica una relación positiva (cuando aumenta x, aumenta y), mientras que un coeficiente negativo indica una relación negativa (cuando aumenta x, disminuye y). La magnitud del coeficiente indica cuánto cambia la variable dependiente por cada cambio unitario en la variable independiente, manteniendo todas las demás variables constantes.

Finalmente, se evalúa la calidad del modelo utilizando métricas como el error cuadrático medio (MSE) o el coeficiente de determinación (R^2), que indican qué tan bien se ajusta el modelo a los datos. Para aplicar este método se usará el paquete de R lm (Weisberg, 2019).

Además, se aplicará también, como se ha dicho, el algoritmo de Random Forest al modelo, este es un poderoso algoritmo de aprendizaje supervisado que se utiliza tanto para problemas de regresión como de clasificación en Machine Learning. Funciona mediante la combinación de múltiples árboles de decisión individuales en un "ensamble" para obtener predicciones más precisas y robustas.

En este proceso se generan múltiples muestras aleatorias con reemplazo del conjunto de datos original, y cada muestra se utiliza para entrenar un árbol de decisión

independiente. Una vez entrenados los árboles individuales, se combinan para formar el Random Forest mediante un proceso conocido como “bagging”.

Durante la etapa de predicción, cada árbol emite su propia predicción, y para problemas de clasificación, se realiza una votación entre los árboles para determinar la clase final, mientras que, para problemas de regresión, se promedian las predicciones de todos los árboles.

Además del muestreo con reemplazo, el algoritmo introduce aleatoriedad adicional durante la construcción de cada árbol de decisión, considerando solo un subconjunto aleatorio de características en cada división de un nodo. Esto ayuda a diversificar los árboles y reduce la correlación entre ellos, lo que mejora la precisión general del modelo.

El Random Forest es robusto frente al sobreajuste gracias a la combinación de múltiples modelos y es capaz de manejar conjuntos de datos grandes con muchas características y observaciones.

Este algoritmo tiene varios hiperparámetros que afectan su rendimiento, como el número de árboles en el bosque, la profundidad máxima de los árboles y el número de características consideradas en cada división.

Una de las ventajas del Random Forest es que además de generalmente brindar buenos rendimientos en las predicciones, este algoritmo también permite estudiar la importancia de las variables, útil para el análisis de datos y la selección de características. Utiliza los índices de Mean Decrease Accuracy (MDA), el cual expresa cuánta precisión pierde el modelo al aleatorizar cada variable y el coeficiente de Gini, que mide cómo cada variable contribuye a la homogeneidad de los nodos y hojas en el bosque aleatorio resultante. Cuanto mayor sea el valor de MDA o la puntuación de Gini, mayor será la importancia de la variable en el modelo. Para aplicar el algoritmo en este modelo se utilizó el paquete de R llamado randomForest , Liaw y Wiener (2002)

4. Resultados

En este capítulo se encuentran y se interpretan los resultados obtenidos durante el análisis del balance competitivo en las cinco grandes ligas europeas, mediante el uso de índices estadísticos elaborados mediante los datos obtenidos con el paquete WorldFootballR (Zivkovic, 2024) y las variables económicas introducidas en el modelo de *Machine Learning*. Para interpretar el resultado de los índices se ha revisado la clasificación de los equipos para las temporadas resultantes en la web de BeSoccer (https://www.besoccer.com/).

4.1. Análisis índices Balance Competitivo

El análisis de los índices de balance competitivo en las ligas de fútbol europeas proporciona una visión integral de la dinámica de la competencia en este deporte. En este estudio se explorará la tarea de interpretar los resultados de estos índices, uno por uno, con el fin de desentrañar los matices y las tendencias que subyacen en la competitividad de las ligas.

El balance competitivo es un componente esencial para garantizar la emoción y la equidad en el fútbol, y los índices diseñados para medirlo ofrecen valiosas perspectivas sobre la distribución de poder entre los equipos, la variabilidad en el rendimiento y la justicia en los resultados. Sin embargo, interpretar adecuadamente estos índices requiere un análisis cuidadoso y detallado.

4.1.1. SHICB

Al ser el SHICB una versión estandarizada del HICB, se ha optado por no incluir este último en este análisis. Más adelante en el punto 4.1.6 se visualizará lo correlacionados que están ambos índices.

La Tabla 2 muestra los resultados del SHICB para las cinco grandes ligas europeas analizadas. Como se ha explicado, este corresponde a una versión normalizada del HICB. Con esto se obtiene que la cantidad de equipos en una liga y temporada determinada no afecte al resultado final. Durante este análisis se ha abarcado desde la temporada 2000 hasta la 2023.

Para este índice un valor de cien indica una liga desequilibrada competitivamente, cuanto menor sea el resultado, menor desequilibrio existe.

En primer lugar, se procederá a analizar el resultado de La Liga en España, se encuentra que el desbalance ha ido aumentado con los años, aunque con cierta tendencia a disminuir los últimos años. Se observa que la temporada con mayor desbalance competitivo es la temporada 2015, con un resultado de 84.2473 en el HICB, observando la clasificación se ve como los dos primeros clasificados, Barcelona y Real Madrid obtuvieron 30 victorias y 94 y 92 puntos respectivamente, ambos frente a las 23 victorias y del tercer clasificado, Atlético de Madrid, con 78 puntos.

La siguiente temporada más desbalanceada se halla en la 2017, con un resultado del SHICB de 83.7718, bastante próximo del anterior resultado. En esta temporada ocurre algo similar a la anterior, en este caso el Real Madrid y Barcelona con 93 y 90 dominan

la liga con 29 y 28 victorias respectivamente, frente al tercer clasificado, el Atlético de Madrid con 23 victorias con 78 puntos.

En último lugar la temporada que muestra un equilibrio competitivo más balanceado se encuentra en el año 2000, con un resultado de 75.7749. En este caso se comentarán los resultados de los 5 primeros clasificados, puesto a la proximidad de resultados. El campeón fue el Deportivo de La Coruña con 69 puntos y 38 victorias, seguido del Barcelona con 64 puntos y 19 victorias, el tercer clasificado fue el Valencia con 64 puntos y 18 victorias, el Real Zaragoza con 63 puntos y 16 victorias, final y quinto clasificado fue el Real Madrid con 62 puntos y 16 victorias.

A continuación, se analizará el resultado del SHICB en la Premier League en Inglaterra, existe un aumento del desequilibrio competitivo a lo largo de las temporadas, lo que propicia que ciertos equipos tienen un dominio competitivo mayor que otros. La temporada más desbalanceada se encuentra en la 2019, con un resultado de 83.8009, observando la clasificación se ve como los dos primeros clasificados, Manchester City y Liverpool obtuvieron 32 y 30 victorias junto a 98 y 97 puntos respectivamente, ambos frente a las 21 victorias y del tercer clasificado, Chelsea con 72 puntos.

La siguiente temporada más desbalanceada se halla en la 2008, con un resultado del SHICB de 83.0764, bastante próximo del anterior resultado. En esta temporada los dos mejores equipos de la competición son Manchester United y Chelsea con 87 y 85 puntos. Dominan la liga con 27 y 25 victorias respectivamente, frente al tercer clasificado el arsenal con 24 victorias y 83 puntos destaca sobre todo el último clasificado, el Derby Country con 1 victoria y 11 puntos, existiendo una diferencia de 80 puntos entre primero y último, incluso de 24 puntos entre penúltimo y último.

En último lugar la temporada que muestra un equilibrio competitivo más balanceado se encuentra en 2011, con un resultado de 77.3600. En esta temporada se observa mayor equilibrio competitivo, en este caso se comentarán los resultados de los 4 primeros clasificados, puesto a la proximidad de resultados. El campeón fue el Manchester United con 80 puntos y 23 victorias, seguido del Chelsea con 71 puntos y 21 victorias, el tercer clasificado fue el Manchester City con 71 puntos y 21 victorias y Arsenal con 68 puntos y 19 victorias. En esta temporada el equilibrio competitivo se encuentra en la mitad de la tabla donde desde el octavo puesto al diecisieteavo solo existe una diferencia de 9 puntos.

En cuanto a la Bundesliga en Alemania, el SHICB presenta una evolución a lo largo de las temporadas, lo que indica que existe cierto desequilibrio competitivo y por tanto existen ciertos equipos que presentan un dominio competitivo mayor que otros. Destaca la temporada 2014 como en la que mayor desequilibrio competitivo existe, con un resultado de 82.0753 en el SHICB, observando la clasificación se ve como los dos primeros clasificados, Bayern Munich y Borussia Dortmund obtuvieron 29 y 22 victorias junto a 90 y 71 puntos respectivamente, ambos frente a las 19 victorias y del tercer clasificado con 64 puntos, Schalke 04.

Seguida de esta, se presenta la temporada 2019, con un resultado de 81.8879 en el SHICB, observando la clasificación se ve como los dos primeros clasificados, Bayern Munich y Borussia Dortmund obtuvieron 24 y 23 victorias junto a 78 y 76 puntos respectivamente, ambos frente a las 19 victorias y del tercer clasificado con 66 puntos, RB Leipzig.

Por último, la más equilibrada competitivamente durante este análisis, se encuentra en la temporada 2001, con un resultado de 76.1075. En esta temporada se observa mayor equilibrio competitivo, en este caso se comentarán los resultados de los 5 primeros clasificados, puesto a la proximidad de resultados. El campeón fue el Bayern Munich con 63 puntos y 19 victorias, seguido del Schalke 04 con 62 puntos y 18 victorias, el tercer clasificado fue el Borussia Dortmund con 58 puntos y 16 victorias, el Bayern Leverkusen con 58 puntos y 26 victorias, final y quinto clasificado fue el Hertha BSC con 57 puntos y 17 victorias.

Siguiendo con la Serie A en Italia, el SHICB presenta una evolución a lo largo de las temporadas, aunque bastante estable, pese a esto indica que existe cierto desequilibrio competitivo y por tanto existen ciertos equipos que presentan un dominio competitivo mayor que otros. Destaca la temporada 2021 como la que mayor desequilibrio competitivo existe, con un SHICB de 83.9187. Inter de Milán y Milán, los dos primeros clasificados, acumularon 28 y 24 victorias respectivamente, junto a 91 y 78 puntos, en este caso esta temporada existió un dominio del Inter de Milán sobre los demás equipos, ya que el tercer clasificado fue el Atalanta con 78 puntos y 23 victorias, el cuarto la Juventus con 23 victorias y 78 puntos también y el quinto el Napoli con 77 puntos y 24 victorias.

Seguida de esta, se presenta la temporada 2018, con un SHICB de 83.5743 muy similar al anterior, destaca la similitud a la temporada 2017 muy pareja con esta en cuanto a resultado. Juventus y Napoli dominaron la liga con 95 y 91 puntos, y 30 y 28 victorias respectivamente, dejando atrás al tercer clasificado, Roma, con 23 victorias y 77 puntos.

La más equilibrada competitivamente durante este análisis, se encuentra en la temporada 2005, con un resultado de 78.2944, evidenciando un mayor equilibrio competitivo que las demás temporadas. En este caso el primer clasificado fue la Juventus con 26 victorias y 86 puntos, pero en esta temporada lo interesante se observa en la media tabla, donde desde el octavo clasificado, la Roma con 11 partidos ganados y 45 puntos, solo existe una diferencia de 10 puntos con el último clasificado, la Atalanta con 8 partidos ganados y 35 puntos.

Por último, la Ligue 1 en Francia, el SHICB presenta una evolución a lo largo de las temporadas, aumentando más los últimos diez años. Esto indica que existe cierto desequilibrio competitivo y, por tanto, existen ciertos equipos que presentan un dominio competitivo mayor que otros. Destaca la temporada 2023 como la que mayor desequilibrio competitivo existe, con un SHICB de 82.1056. PSG y Lens, son los dominantes de esta temporada, con 85 y 84 puntos junto a 27 y 25 victorias respectivamente, el tercer clasificado fue el Olympique de Marseille con 73 puntos y 22 victorias.

Seguida de esta, se presenta la temporada 2018, con un SHICB de 80.9686, destaca la similitud a la temporada 2017 muy pareja con esta en cuanto a resultado. El PSG dominó la liga con 29 victorias y 93 puntos, seguido de los siguientes clasificados el Mónaco con 80 puntos y 24 victorias, Olympique Lyonnais con 78 puntos y 23 victorias y finalmente el Olympique Marseille con 77 puntos y 22 victorias.

La más equilibrada competitivamente durante este análisis se encuentra en la temporada 2000, con un resultado de 74.9383. En este caso el primer clasificado fue el Mónaco con

20 victorias y 65 puntos, pero en esta temporada lo interesante se observa desde el quinto clasificado al dieciseisavo, siendo el Lens quinto con 14 victorias y 49 puntos y dieciseisavo el Nancy con 11 victorias y 42 puntos, solamente 7 puntos entre estas posiciones, evidenciando una alta competitividad en esta parte de la tabla.

Temporada	España	Inglaterra	Alemania	Italia	Francia
2000	75,7749	79,6012	78,4158	80,1989	74,9383
2001	77,0811	78,0984	76,1075	79,4396	76,2283
2002	75,9836	80,2402	79,8619	80,0632	76,3906
2003	77,2925	78,9907	76,7317	79,4420	76,8474
2004	77,0516	79,0832	79,1878	82,5707	78,0926
2005	78,3712	80,7159	78,7875	78,2944	75,9751
2006	78,6568	81,5884	79,5183	81,6248	78,0050
2007	77,7153	79,5029	77,5860	80,6722	76,0224
2008	78,1603	83,0764	78,2508	79,5062	77,4218
2009	78,3155	81,5660	79,5102	79,2358	78,7871
2010	81,8724	82,3106	78,6110	78,6735	78,8978
2011	80,0137	77,3600	77,6623	78,8415	76,8321
2012	80,1918	80,7813	80,1309	78,8958	78,5343
2013	80,9213	81,3558	80,9829	80,6777	77,8091
2014	81,4366	82,2229	82,0753	82,6044	79,8545
2015	84,2473	79,8480	78,7462	80,0879	78,6193
2016	81,3631	79,2795	80,4743	80,6411	79,0366
2017	83,7718	82,8899	78,9395	83,5247	80,8784
2018	81,3814	82,4947	79,1136	83,5743	80,9686
2019	78,1398	83,8009	81,8879	81,9150	79,8783
2020	79,8220	81,0650	80,9074	81,5173	79,6619
2021	81,3913	80,1652	80,1250	83,9187	80,0763
2022	79,1682	82,4676	79,7244	81,7939	79,5604
2023	79,4187	81,3633	78,9036	81,5508	82,1056

Tabla 2. SHICB de las cinco grandes ligas europeas de la temporada 2000 hasta la 2023

Fuente: Elaboración propia con los datos de Fbref y R.

4.1.2. Ratio de desviación estándar

La Tabla 3 muestra los resultados del Ratio de Desviación estándar o RSD para las cinco grandes ligas europeas analizadas desde la temporada 2000 hasta la 2023. Un resultado del RSD bajo indica un mayor equilibrio competitivo en la liga y temporada analizada.

En primer lugar, se procederá a analizar el resultado de La Liga en España, se encuentra que la ratio ha ido aumentando con los años, aunque con cierta tendencia a aumentar los últimos años siete años. Se observa que la temporada con mayor desbalance competitivo es en la temporada 2019, siendo la más desbalanceada con un resultado de 235.2520 en el RSD, observando la clasificación se ve como los tres primeros clasificados, Barcelona, Atlético de Madrid y Real Madrid obtuvieron 26, 22 y 21 victorias respectivamente, frente a las 15 victorias del cuarto clasificado, Valencia.

La siguiente temporada más desbalanceada corresponde a la 2018 con un resultado del RSD de 229.2889. En esta temporada ocurre algo similar a la anterior, en este caso el Real Madrid, Barcelona, Atlético de Madrid y Valencia con 28, 23 y 22 victorias respectivamente, frente al quinto clasificado el Villareal con 18 victorias.

En cuanto a la temporada con menor RSD, esta se remonta a 2011 con un resultado de 147.1709. En esta temporada se observa mayor equilibrio competitivo, en este caso se comentan por una parte los resultados desde el Valencia, tercer clasificado con 21 victorias hasta el Atlético de Madrid con 17 victorias, habiendo un lapso de solamente 4 victorias. Por otra parte, la media tabla donde existe una diferencia con el octavo clasificado, Espanyol hasta el diecisieteavo Mallorca una diferencia de 3 victorias, con 15 y 12 respectivamente.

A continuación, se procederá a analizar el resultado de Premier League en Inglaterra, se encuentra que la ratio ha ido aumentando con los años, aunque con cierta tendencia a disminuir los últimos años. Se observa que la temporada con mayor desbalance competitivo es en la temporada 2015, siendo la más desbalanceada con un resultado de 244.7170 en el RSD, observando la clasificación se ve como los tres primeros clasificados dominaron esta campaña, Chelsea, Manchester City y Arsenal obtuvieron 26, 24 y 22 victorias respectivamente, frente a las 20 victorias del cuarto clasificado, Manchester United y la lucha más justa en la clasificación de la media tabla.

La siguiente temporada más desbalanceada corresponde a la 2017 con un resultado del RSD de 233. 1228. En esta temporada domina el Chelsea y Tottenham Hotspur con 30 y 26 victorias, mientras que desde el octavo clasificado al diecisieteavo solo existe una diferencia de una victoria.

En cuanto a la temporada con menor RSD, esta se remonta a 2002 con un resultado de 116.4931. En esta temporada se observa mayor equilibrio competitivo, en este caso no existe una diferencia abrumadora en la clasificación como en las otras temporadas, ya que el Arsenal es el campeón con 26 victorias, le siguen Liverpool y Manchester United con 24. Esta es una temporada donde las victorias se reparten bastante equitativamente y no hay un dominador como tal.

Le sigue a esta, el resultado de la Bundesliga en Alemania se encuentra que la ratio ha ido aumentando con los años, aunque presenta cierta estabilidad. Se observa que la temporada con mayor desbalance competitivo es en la temporada 2014, siendo la más

desbalanceada con un resultado de 216.3626 en el RSD, observando la clasificación se ve como domina esta temporada el Bayern Munich la liga con 29 victorias y campeón frente al segundo calificado, el Borussia Dortmund con 22 victorias, mientras que los otros en la clasificación están compitiendo constantemente.

La siguiente temporada más desbalanceada corresponde a la 2019 con un resultado del RSD de 201. 1979. En esta temporada domina el Bayern de Munich y Borussia Dortmund con 24 y 23 victorias, frente al tercer clasificado el RB Leipzig con 19 victorias.

En cuanto a la temporada con menor RSD, esta se remonta a 2001 con un resultado de 124.0884. En esta temporada se observa mayor equilibrio competitivo, en este caso no existe una diferencia abrumadora en la clasificación como en las otras temporadas, ya que el Bayern Munich es el campeón con 19 victorias, le siguen Schalke 04 y Borussia Dortmund con 18 y 16 victorias respectivamente, esta es una temporada donde las victorias se reparten bastante equitativamente y no hay un dominador como tal.

Siguiendo, se analizará el resultado de la Serie A en Italia, se encuentra que la ratio ha ido aumentando con los años, aunque con cierta tendencia a aumentar las últimas diez temporadas. Se observa que la temporada con mayor desbalance competitivo es en la temporada 2021, siendo la más desbalanceada con un resultado de 240.8664 en el RSD. Observando la clasificación se ve como existe un claro dominador, el campeón de esta temporada, el Inter de Milán con 28 victorias, mientras que desde el segundo clasificado al sexto se encuentra cierta competitividad, igual que a media tabla.

La siguiente temporada más desbalanceada corresponde a la 2017 con un resultado del RSD de 233. 2476. En esta temporada domina la Juventus, Roma y Napoli con 29,28 y 26 victorias.

En cuanto a la temporada con menor RSD, esta se remonta a 2005 con un resultado de 158,9003. En esta temporada se observa mayor equilibrio competitivo, en este caso existe cierta diferencia en la clasificación entre el primero, la Juventus con 26 victorias y el Milán con 23 victorias. En esta temporada se observa que existe una competitividad elevada en la clasificación debido a la diferencia de solo 2 victorias en el lapso entre sexto clasificado y diecinueveavo.

Finalmente, se procederá a analizar el resultado de la Ligue 1 en Francia, se encuentra que la ratio ha ido aumentando con los años, con cierta tendencia a aumentar bastante. Se observa que la temporada con mayor desbalance competitivo es en la temporada 2023, siendo la más desbalanceada con un resultado de 217.1077 en el RSD. Observando la clasificación se observa como dominan la liga PSG y Lens con 27 y 25 victorias, mientras que en los demás equipos a mitad tabla sí existe cierta competitividad entre ellos.

La siguiente temporada más desbalanceada corresponde a la 2017 con un resultado del RSD de 200.5809. En esta temporada domina absolutamente el Mónaco y PSG con 30 y 27 victorias, frente al tercero el Nice con 22 victorias

En cuanto a la temporada con menor RSD, esta se remonta a 2000 con un resultado de 106.0275. En esta temporada se observa mayor equilibrio competitivo, en este caso el Mónaco es el campeón con 20 victorias, mientras que el segundo clasificado el PSG

obtuvo 16 victorias, destacando que del segundo al catorceavo existe una diferencia de 5 victorias.

Temporada	España	Inglaterra	Alemania	Italia	Francia
2000	177,5175	123,8353	157,5109	176,6883	106,0275
2001	162,2129	144,8947	124,0884	165,9153	119,4792
2002	191,0679	116,4931	182,4689	162,2719	122,5765
2003	167,2658	145,0475	131,9566	169,6277	133,1901
2004	173,7321	146,1889	155,6324	212,7428	158,0613
2005	202,3409	161,0217	172,5602	158,9003	115,0214
2006	207,4178	175,1533	177,6432	223,3961	151,4107
2007	181,6819	156,4495	157,4621	199,4730	130,0330
2008	212,6212	167,1912	149,2774	180,9179	143,7912
2009	202,0052	168,8399	178,7650	182,5868	159,7782
2010	208,8353	228,4174	146,2593	161,3310	163,1664
2011	147,1709	196,2177	150,3169	174,6465	131,2628
2012	203,9160	198,3589	194,4343	162,5541	169,5358
2013	206,6485	202,9150	195,8037	191,3938	147,8938
2014	218,5066	208,4370	216,3626	221,1779	179,0712
2015	187,2013	244,7170	170,5770	177,7047	153,0484
2016	160,4271	216,8523	193,4828	197,1753	175,9501
2017	218,6841	233,1228	155,5830	233,2476	200,5809
2018	229,2889	196,6760	176,5250	218,6017	195,9705
2019	235,2520	158,0701	201,1979	212,7254	193,0941
2020	206,1116	180,4503	195,2925	200,7948	163,5595
2021	190,4289	214,3017	177,3507	240,8664	187,0829
2022	219,3923	172,4759	186,9338	210,7630	185,6038
2023	204,7294	188,7339	166,4820	221,8032	217,1077

Tabla 3. RSD de las cinco grandes ligas europeas de la temporada 2000 hasta la 2023

Fuente: Elaboración propia con los datos de Fbref y R

4.1.3. Medidas de porcentaje de empate

La Tabla 4 muestra los resultados del porcentaje de empate o Draw% para las cinco grandes ligas europeas analizadas desde la temporada 2000 hasta la 2023. Un resultado de Draw% mayor, implica que los equipos están muy igualados entre sí y por tanto existe una mayor presencia de equilibrio competitivo en una liga.

En primer lugar, se procederá a analizar el resultado de La Liga en España, se encuentra que el porcentaje ha ido disminuyendo con los años, aunque con cierta tendencia a aumentar las últimas temporadas. Se observa que la temporada con mayor desbalance competitivo es la temporada 2011, siendo la más desbalanceada con un resultado de 0.2079 en el Draw%. Observando la clasificación se ve como resulta en una temporada muy apretada, sobre todo desde el octavo clasificado, el Espanyol con 48 puntos y el decimoctavo el RC Deportivo con 43 puntos, existiendo una diferencia de solamente 5

puntos entre estos, aunque esto no ocurre con los clasificados arriba, donde si existe esa desigualdad con el Barcelona y Real Madrid en cabeza con 96 y 92 puntos, respecto el tercero Valencia con 71.

La siguiente temporada más desbalanceada corresponde a la 2009 con un resultado del Draw% de 0.2184. En esta temporada ocurre algo similar a la anterior, existe un dominio absoluto del Barcelona con 87 puntos, frente al Real Madrid con 78 y Sevilla como tercero con 70 puntos, realmente la competitividad se encuentra a mitad tabla con una diferencia entre el onceavo clasificado y el decimoctavo de 4 puntos.

En cuanto a la temporada con mayor Draw%, esta se remonta a 2000 con un resultado de 0.2947. En esta temporada se observa mayor equilibrio competitivo, igual que ocurre con el SHICB, esta temporada destaca por la cercanía de puntos entre todos los clubs competidores.

A continuación, se procederá a analizar el resultado de Premier League en Inglaterra, se encuentra que el porcentaje es bastante estable, como temporada más desbalanceada destaca la 2019 con un resultado de 0.1868, este resultado induce en un desequilibrio competitivo grande y observando la clasificación se aprecia como Manchester City y Liverpool dominaron la liga con 98 y 97 puntos respectivamente, frente al tercer clasificado el Chelsea con 72 puntos, además no existe cierta cercanía de resultados entre los demás equipos.

La siguiente temporada más desbalanceada corresponde a la 2006 con un resultado del Draw% de 0.2026. En esta temporada ocurre algo similar a la anterior, existe un dominio del Chelsea con 91 puntos, Manchester United con 83 y Liverpool como tercero con 82 puntos frente al cuarto clasificado Arsenal con 67 puntos.

En cuanto a la temporada con mayor Draw%, esta se remonta a 2011 con un resultado de 0.2921. En esta temporada se observa mayor equilibrio competitivo, siendo esto por la cercanía de resultados en la clasificación para la parte alta, media y baja.

Siguiendo con el análisis, el resultado de la Bundesliga en Alemania se encuentra que el porcentaje es también estable, con tendencia a disminuir. Como temporada más desbalanceada destaca la 2011 con un resultado de 0.2059, observando la clasificación se aprecia como Borussia Dortmund fue campeón con 75 puntos, Bayern Leverkusen le sigue con 68 puntos y Bayern Munich con 65 puntos, se observa cierta distancia entre los tres primeros clasificados, observando la clasificación existe una tendencia a más victorias y menos empates.

La siguiente temporada más desbalanceada corresponde a la 2014 con un resultado del Draw% de 0.2092. En este caso se debe al dominio del Bayern Munich con 90 puntos y poca cercanía de puntos en la tabla en la parte alta de esta.

En cuanto a la temporada con mayor Draw%, esta se remonta a 2006 con un resultado de 0.3137. En esta temporada se observa mayor equilibrio competitivo, siendo esto por la cercanía de resultados en la clasificación, sobre todo desde el sexto clasificado hasta la parte baja de esta.

A continuación, se analizará la Serie A en Italia, se encuentra que el porcentaje es elevado, aunque ha ido decreciendo con el paso de las temporadas, fruto a priori de una

perdida de equilibrio competitivo. Como temporada más desbalanceada destaca la 2017 con un resultado de 0.2105, observando la clasificación se aprecia como la Juventus es campeón con 91 puntos frente al segundo clasificado, la Roma con 87 puntos y tercer clasificado Napoli con 86 puntos dominan la liga, el cuarto clasificado es la Atalanta con 72 puntos, existiendo cierta distancia de puntos entre estos.

La siguiente temporada más desbalanceada corresponde a la 2018 con un resultado del Draw% similar al anterior, siendo este de 0.2184. Se observa como en este caso dominan la liga Juventus y Napoli con 95 y 91 puntos respecto a, por ejemplo, el tercer clasificado, la Roma con 77 puntos.

En cuanto a la temporada con mayor Draw%, esta se remonta a 2005 con un resultado de 0.3289. En esta temporada se observa mayor equilibrio competitivo, siendo esto por la cercanía de resultados en la clasificación además de la gran cantidad de empates que existen entre todos los clubes.

Por último, se analizará la Ligue 1 en Francia, donde se encuentra que el porcentaje es estable, aunque ha ido decreciendo con el paso de las temporadas, fruto a priori de una perdida de equilibrio competitivo. Como temporada más desbalanceada destaca la 2015 con un resultado de 0.2315, se debe principalmente a la poca cercanía de puntos entre los equipos y dominio del PSG con 83 puntos.

La siguiente temporada más desbalanceada corresponde a la 2023 con un resultado del Draw% de 0.2421. Se observa un dominio en esta temporada del campeón PSG con 85 puntos y Lens con 84, además de no existir demasiados empates.

En cuanto a la temporada con mayor Draw%, esta se remonta a 2005 con un resultado de 0.3474. En esta temporada se observa mayor equilibrio competitivo, siendo esto por la cercanía de resultados en la clasificación además de la gran cantidad de empates que existen entre todos los clubes.

Temporada	España	Inglaterra	Alemania	Italia	Francia
2000	0,2947	0,2421	0,2843	0,3072	0,2614
2001	0,2605	0,2658	0,2255	0,2876	0,2712
2002	0,2658	0,2658	0,2222	0,2843	0,2810
2003	0,2763	0,2368	0,2516	0,3007	0,2789
2004	0,2500	0,2842	0,2353	0,2941	0,2605
2005	0,2632	0,2895	0,2124	0,3289	0,3474
2006	0,2763	0,2026	0,3137	0,2842	0,3105
2007	0,2579	0,2579	0,2582	0,3000	0,3053
2008	0,2289	0,2632	0,2549	0,2947	0,3053
2009	0,2184	0,2553	0,2418	0,2500	0,2947
2010	0,2500	0,2526	0,2810	0,2684	0,2553
2011	0,2079	0,2921	0,2059	0,2553	0,3421
2012	0,2474	0,2447	0,2582	0,2921	0,2842
2013	0,2211	0,2842	0,2549	0,2526	0,2842
2014	0,2263	0,2053	0,2092	0,2368	0,2842
2015	0,2395	0,2447	0,2680	0,3158	0,2316
2016	0,2421	0,2816	0,2320	0,2500	0,2842
2017	0,2342	0,2211	0,2418	0,2105	0,2474
2018	0,2263	0,2605	0,2712	0,2184	0,2526
2019	0,2895	0,1868	0,2386	0,2842	0,2895
2020	0,2763	0,2421	0,2222	0,2237	0,2593
2021	0,2868	0,2184	0,2647	0,2526	0,2500
2022	0,2921	0,2316	0,2386	0,2579	0,2684
2023	0,2342	0,2289	0,2451	0,2632	0,2421

Tabla 4. Draw% de las cinco grandes ligas europeas de la temporada 2000 hasta la 2023

Fuente: Elaboración propia con los datos de Fbref y R

4.1.4. Medida de la dominancia

Se ha realizado un gráfico a modo de mostrar de manera más visual el dominio que existe en cada liga, según la cantidad de veces que un mismo equipo ha ganado la liga, representado por el eje y, por otra parte, el eje x muestra los diferentes equipos que han ganado esa liga durante las temporadas analizadas en este TFG.

Se observa en la Figura 2, la dominancia de campeonatos ganados por liga/país, se observa como la liga con mayor equilibrio competitivo es la Ligue 1 en Francia con 8 campeones distintos y siendo el mismo equipo campeón en 9 ocasiones durante todas las temporadas que se analizan.

El caso contrario, es decir la liga más desequilibrada es la Bundesliga en Alemania, con 5 campeones distintos y siendo 18 veces el mismo equipo campeón durante todas las temporadas que se analizan.

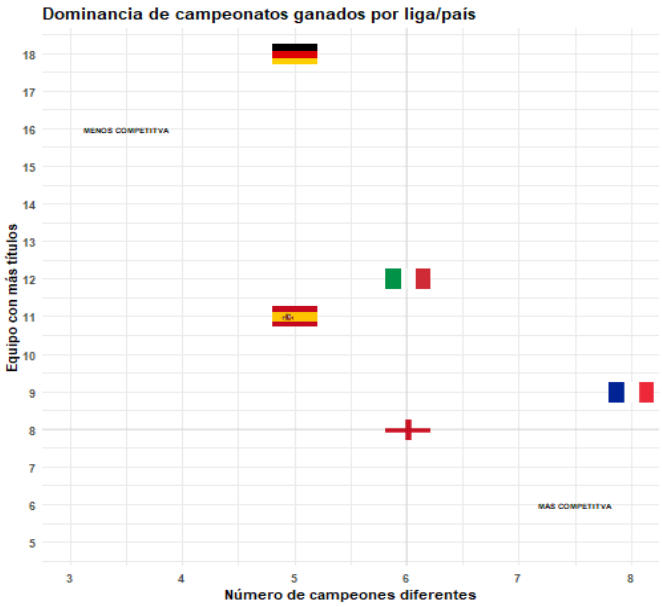

Figura 2. Dominancia de campeonatos ganados por liga/país

Fuente: Elaboración propia con los datos de Fbref y R Studio

4.1.5. Análisis de la evolución de los índices

Se han elaborado unos gráficos que muestran la evolución de las ligas durante las temporadas analizadas y los índices elaborados.

La Figura 3, muestra la evolución de los índices de concentración SHIBC y RSD a lo largo del tiempo, desde el año 2000 hasta el 2023. La Ligue 1 destaca con una tendencia creciente constante en ambos índices, lo que sugiere un aumento en la concentración en esta liga. Por otro lado, la Bundesliga, La Liga, Premier League y Serie A presentan fluctuaciones más pronunciadas. Aunque estas ligas muestran una tendencia general decreciente en las últimas temporadas para el índice SHIBC, la tendencia no es tan marcada para el RSD. Es importante notar que la variabilidad de los datos, indicada por el área sombreada gris, sugiere que hay temporadas con comportamientos atípicos que podrían influir en la interpretación de las tendencias a largo plazo.

La Figura 4 presenta el indicador de Draw% a lo largo de las temporadas, desde el año 2000 hasta el 2023. Se observa que la Bundesliga mantiene una estabilidad en los valores bajos, lo que indica una tendencia decreciente en el porcentaje de empates. Este comportamiento sugiere un menor balance competitivo en comparación con las otras ligas. En contraste, la Ligue 1 muestra una menor frecuencia de empates, pero tendencia creciente, lo cual podría interpretarse como un signo hacia mayor balance competitivo La Premier League, La Liga y la Serie A exhiben fluctuaciones en sus porcentajes de empate, pero sin una tendencia clara que indique un cambio significativo en la competitividad a lo largo del tiempo. Estos hallazgos concuerdan con los resultados

previos de Ramchandani (2018) aunque es notable la evolución de la Bundesliga hacia una menor competitividad los últimos años.

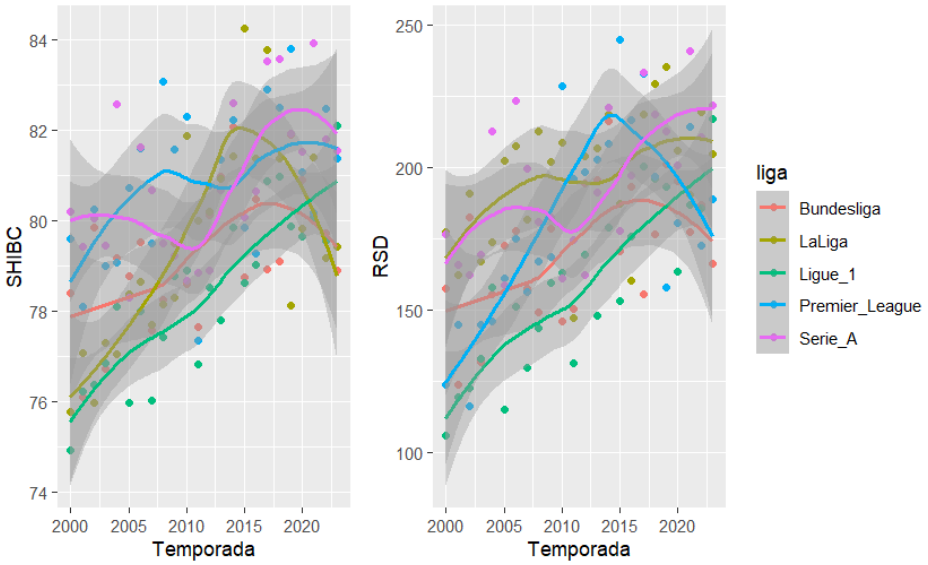

Figura 3. Evolución de los índices SHIBC y RSD a lo largo de las temporadas para cada liga/país

Fuente: Elaboración propia con los datos de Fbref y R

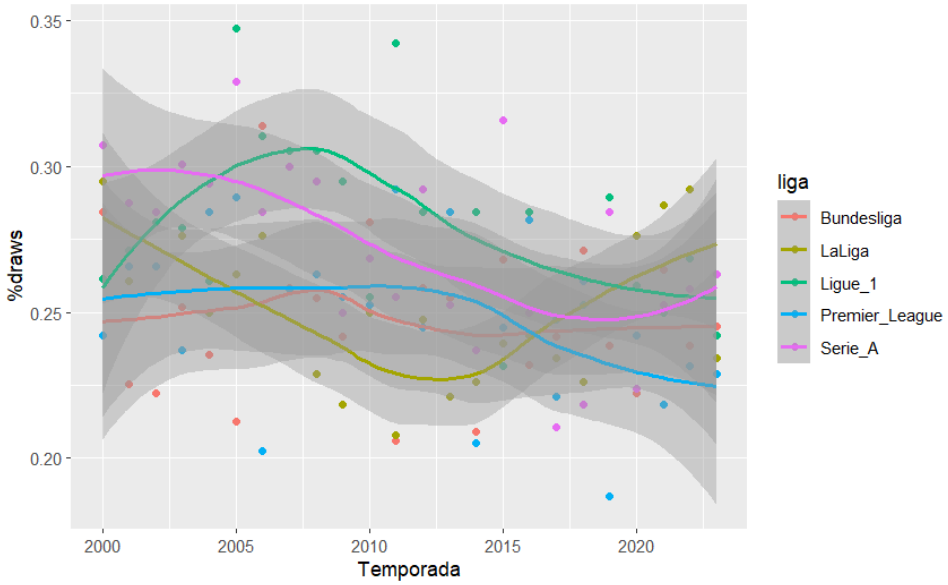

Figura 4. Evolución de Draw% a lo largo de las temporadas para cada liga/país

Fuente: Elaboración propia con los datos de Fbref y R

4.1.6. Análisis de correlación de los índices

Se han elaborado unos gráficos que muestran la correlación entre los índices elaborados.

En la Figura 5 se encuentran las correlaciones de los índices. En primer lugar, se observa una interrelación muy fuerte entre HICB y SHICB, con un resultado de 0.9986, lo que indica que los índices están muy relacionados, dado que el SHICB es una versión estandarizada del HICB, se usará este para los análisis.

En segundo lugar, se encuentra una correlación positiva moderada, pese a que miden aspectos diferentes, este es de 0.5724 entre el SHICB y RSD.

Seguidamente se observa una correlación negativa de -0.5113 entre SHICB y el Draw%, lo indica que a medida que aumenta el SHICB, la proporción de partidos empatados tiende a disminuir.

Finalmente se analiza la correlación entre el RSD y el Draw%, el cual presenta una correlación negativa de -0.3474, lo indica que a medida que aumenta el RSD, la proporción de partidos empatados tiende a disminuir.

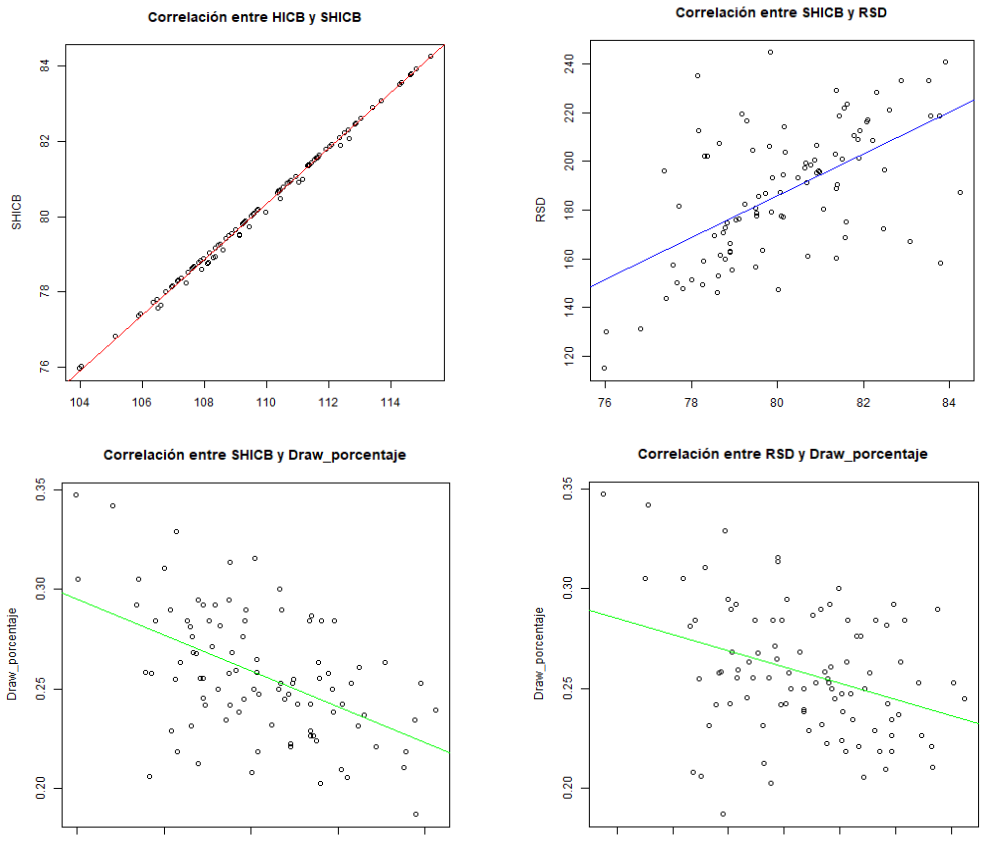

Figura 5. Correlación de los índices del Balance Competitivo

Fuente: Elaboración propia con los datos de Fbref y R

4.2. Análisis de los modelos de *Machine Learning*

A continuación, se va a examinar la relación entre índices diseñados para medir el desequilibrio competitivo y variables económicas relevantes. Estos índices han sido desarrollados para capturar diversos aspectos de la competencia en las cinco grandes ligas. Se emplearán técnicas de *Machine Learning* para analizar cómo estas variables interactúan entre sí y si afectan al desbalance competitivo. El objetivo es proporcionar una comprensión más profunda de cómo las fuerzas económicas impactan en la dinámica competitiva.

Durante el análisis se usará regresión lineal y el algoritmo Random Forest, utilizando en R las librerías car (Weisberg, 2019) y randomForest Liaw y Wiener (2002).

Antes de empezar, cabe destacar que los datos para las variables económicas desde la temporada 2000 a la 2004 no se han podido obtener, por lo que se obviarán estas temporadas en el modelo elaborado.

4.2.1. Análisis del método de Regresión Lineal

En el análisis de regresión lineal, la variable categórica **liga**, que incluye las cinco grandes ligas (**ENG**, **ESP**, **FRA**, **GER**, **ITA**), se maneja mediante variables dummies. Dado que hay cinco categorías, se crean cuatro variables dummies. En este caso, la liga **ENG** se utiliza como la categoría de referencia. Esto significa que los coeficientes de las ligas **ESP**, **FRA**, **GER** e **ITA** se interpretan en relación con la liga **ENG**. Aunque **ENG** no aparece explícitamente en el modelo, su influencia está implícita y sirve como punto de comparación para los efectos de las demás ligas sobre los índices o variables dependientes.

En la Tabla 5, se observa una salida de R con los resultados del modelo de *Machine Learning para* el índice SHICB como variable dependiente y como explicativas las variables que miden el desequilibrio competitivo y las económicas y a la derecha con transformación logarítmica del índice y de las variables económicas. En el modelo de la Tabla 5 se ha procedido a centrar las variables económicas y a restar la primera temporada, estas transformaciones permiten una mejor estimación e interpretación de los coeficientes del modelo.

A continuación, se procede a interpretar el resultado:

En primer lugar, se observa que el intercepto, que corresponde a la Premier League (**ligaENG),** por ser la categoría de referencia, tiene un valor de 81.71 lo que indica el resultado del índice de la Premier en la primera temporada cuando el valor medio coincide con su media y la desviación típica también. Dado que el p-valor tiene un resultado de prácticamente cero, es estadísticamente significativo, y por tanto se debería de tener en cuanta como un indicador preciso del balance competitivo para la primera temporada y cuando las variables están centradas.

Todas las estimaciones están comparadas en relación con el SHICB de la Premier League.

En cuanto a las estimaciones de los índices elaborados para cada liga:

- En primer lugar, La Liga (**ligaESP**)**,** tiene un coeficiente, negativo -2.169 y un p-valor de 0.000649<0.05 siendo estadísticamente significativa, lo que indica que

existe un menor SHICB en la liga española respecto a la Premier y, por tanto, existe mayor Balance Competitivo en España.

- En cuanto a la Ligue 1 (**ligaFRA**), presenta un coeficiente de -2.8860 y un p-valor prácticamente de cero, lo que sugiere que un menor desequilibrio en la liga francesa respecto a la inglesa, lo que supone a priori que está asociado también a un menor SHICB.
- En cuanto a la Bundesliga (**ligaGER),** el coeficiente negativo de -2.074 y un p-valor de 0.00829<0.05 implica que es estadísticamente significativa y por tanto un menor desequilibrio competitivo en la liga alemana y por tanto mayor balance competitivo, respecto a la Premier.
- Por último, la Serie A (**ligaITA),** con un coeficiente de -0.5609 pero con un p-valor de 0.325142>0.05 indica que no hay evidencia suficiente para afirmar que el desequilibrio competitivo en la liga italiana medido por SHICB es diferente en comparación con la liga inglesa.
- Por tanto, la liga española, francesa y alemana tienen menor SHICB que la inglesa y por ello mayor balance competitivo mientras que la italiana es similar a la inglesa.

Por otro lado, el resultado de temporada, 0.1073 siendo este positivo, pero con un p-valor de 0.077130>0.05 indica que no es estadísticamente significativa y que el paso del tiempo no implica una tendencia hacia un mayor SHICB a lo largo de las temporadas.

En cuanto a las variables económicas, estas muestran que el valor de mercado medio durante las temporadas analizadas (**Valor_Media)** con coeficiente de -1.1373e-08 y la desviación típica de este (**Valor_DesvTip)** con coeficiente de 1.683e-08 y ambas con p-valores 0.008998<0.05 y 0.006153<0.05 respectivamente indican efectos significativos de las variables, aunque fuertemente influenciados por las unidades en que se miden. El primer coeficiente indica que cada euro que aumenta la media del valor de mercado disminuye el índice y, por tanto, tiende a un mayor balance competitivo; en el caso de la desviación típica ocurre lo contrario.

En cuanto al ajuste del modelo los valores de R^2 son de 0.4638 y el ajustado 0.4207, lo que sugiere que el modelo tiene un ajuste moderado de la variabilidad del índice SHICB. Pese que algunas tendencias son significativas e interesantes, hay una variabilidad restante del 57.93% que el modelo no explica, indicando que hay otros efectos de factores explicativos que no captura el modelo.

Finalmente se observa que tanto el estadístico F como el p-valor tiene relevancia estadística y por tanto al menos una de las variables explicativas está relacionada con el SHICB y el modelo es útil.

En el caso de la Tabla 6 se han aplicado logaritmos al SHICB y a las variables económicas para conocer el incremento porcentual del índice por cada aumento porcentual de estas variables. En primer lugar, se observa que las ligas estadísticamente significativas, coinciden con las anteriormente dichas, además con coeficientes negativos que indican que tienen a priori mayor Balance Competitivo y por tanto menor SHICB, que la Premier, menos la italiana.

En cuanto a las variables económicas, ambas son estadísticamente significativas. En el caso de la Media, por cada 1% que aumenta esta, hay una disminución del 3.2984% del índice, indicando por tanto una mejora del Balance Competitivo. En el caso de la

desviación típica, por cada 1% que incrementa esta, hay un aumento del SHICB de 3.5930%, lo que implica un empeoramiento del Balance Competitivo.

Tabla 5. Resultado del modelo de Machine Learning para el índice SHICB mediante regresión lineal

Term	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	81,7076218	0,446794259	182,875	2,7798E-114
ligaESP	-2,168611338	0,612941874	-3,538	0,000649341
ligaFRA	-2,859846751	0,686179216	-4,168	7,23664E-05
ligaGER	-2,073743128	0,598707367	-3,464	0,000828948
ligaITA	-0,56094354	0,566872192	-0,990	0,325142459
temporada_centrado	0,107260271	0,059962551	1,789	0,077129644
Valor_Media_centrado	-1,13733E-08	4,25575E-09	-2,672	0,008989691
Valor_DesvTip_centrado	1,68338E-08	5,99489E-09	2,808	0,006152869
	R.squared	**Adj.R.squared**	**F.statistic**	**p.value**
	0,4638	0,4207	10,7508	1,08884E-09

Fuente: Elaboración propia con R

Tabla 6.Resultado del modelo de Machine Learning aplicando logaritmos para el índice SHICB mediante regresión lineal

Term	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	1,992890	1,848274493	1,078243439	0,28390564
ligaESP	-0,026583	0,008075941	-3,29166307	0,00144015
ligaFRA	-0,033025	0,011853704	-2,786017381	0,00654909
ligaGER	-0,023305	0,009042739	-2,577177744	0,01164541
ligaITA	-0,005965	0,007635827	-0,78113041	0,43684622
temporada	0,001173	0,001030003	1,138371894	0,25809031
log(Valor_Media)	-0,032984	0,015316845	-2,153438756	0,03405073
log(Valor_DesvTip)	0,035930	0,010209779	3,519181321	0,00069105
	R.squared	**Adj.R.squared**	**F.statistic**	**p.value**
	0,4833	0,4417	11,6244	2,40183E-10

Fuente: Elaboración propia con R

En la Tabla 7, se observa una salida de R Studio con los resultados del modelo de *Machine Learning para* el índice Draw% como variable dependiente y como explicativas las variables que miden el desequilibrio competitivo y las económicas. En el modelo de la Tabla 7, al igual que en el modelo para SHICB, se ha procedido a centrar las variables económicas y a restar la primera temporada. Estas transformaciones permiten una mejor estimación e interpretación de los coeficientes del modelo.

Todas las estimaciones están comparadas en relación con el Draw% de la Premier League.

A continuación, se procede a interpretar el resultado:

En primer lugar, se observa que el intercepto que corresponde a la Premier League (**ligaENG**), tiene un valor de 0.2417 se interpreta como índice de la Premier en la primera temporada cuando el valor medio coincide con su media y la desviación típica también. Dado que el p-valor tiene un resultado de prácticamente cero, es estadísticamente significativo, y por tanto se debería de tener en cuenta como un indicador preciso del balance competitivo para la primera temporada y cuando las variables están centradas.

En cuanto a las estimaciones de los índices elaborados para cada liga:

- En primer lugar, en La Liga (**ligaESP**)**,** tiene un coeficiente positivo (0.007572) y un p-valor de 0.5468>0.05 no siendo estadísticamente significativo, lo que significa que no existe suficiente evidencia para determinar que el desequilibrio competitivo en la liga española medido por el Draw% es diferente en comparación con la liga inglesa.
- En cuanto a la Ligue 1 (**ligaFRA**), presenta un coeficiente de 0.04132 y un p-valor de 0.041<0.05, por lo que el coeficiente es significativo y por tanto aumento del índice Draw% en esta liga, respecto a la Premier League, lo que sugiere que un aumento de los empates en esta liga y una mejoría en cuanto a equilibrio competitivo se refiere.
- En la Bundesliga (**ligaGER),** existe un aumento del coeficiente de 0.00735 y un p-valor de 0.05665>0.05 lo que implica que no es estadísticamente significativa la diferencia respecto de la inglesa y por tanto el desequilibrio competitivo en la liga alemana medido por el Draw% no es diferente en comparación con la liga inglesa.
- Por último, la Serie A (**ligaITA),** con un coeficiente de 0.02348 y con un p-valor de 0.0456>0.05 indica que la diferencia es estadísticamente significativa y por tanto sí existe un mayor número de empates y con ello mejora el equilibrio competitivo en esta liga, respecto a la Premier League.
- Por tanto, la liga francesa e italiana tienen mayor Draw% que la inglesa y por ello mayor balance competitivo, mientras que la española y alemana son similares a la inglesa.

Por otro lado, el coeficiente de **temporada** es-0.002155 por tanto negativo, pero con un p-valor de 0.0819>0.05, lo que indica que no es estadísticamente significativa y que el paso del tiempo no implica una tendencia significativa hacia un menor Draw% a lo largo de las temporadas.

En cuanto a las variables económicas, estas muestran que el valor de mercado medio durante las temporadas analizadas (**Valor_Media)** con coeficiente de 4.852e-11 y la

desviación típica de este (**Valor_DesvTip)** con lo coeficiente de -2.400e-11 y ambas con p-valores no significativos siendo estos 0.5781>0.05 y 0.8451>0.05 respectivamente indican que no hay efecto de estas variables en los empates.

En cuanto al ajuste del modelo los valores de R^2 son de 0.2779 y el ajustado de 0.2198, lo que sugiere que el modelo tiene un ajuste limitado de la variabilidad del índice Draw%, hay una variabilidad del 78.02% que el modelo no explica completamente.

Finalmente se observa que tanto el estadístico F como el p-valor tienen relevancia estadística y por tanto al menos una de las variables explicativas está relacionada con el Draw%. Pese a esto sigue siendo un modelo pobre que no explica suficientemente el tema de a qué se deben las tendencias en el Draw%.

En el caso de la Tabla 8, se han aplicado logaritmos al Draw% y a las variables económicas para conocer el incremento porcentual del índice por cada aumento porcentual de estas variables. En primer lugar, se observa que, de las ligas estadísticamente significativas, solo lo es la francesa, con coeficiente positivo indica que existe un mayor equilibrio competitivo y mayor Draw% que la Premier.

En cuanto a las variables económicas, solo la desviación típica es estadísticamente significativa, esta muestra que por cada 1% que disminuye esta, hay un decrecimiento del Draw% de -11.2228%, lo que implica un empeoramiento del Balance Competitivo.

Tabla 7. Resultado del modelo de Machine Learning para el índice Draw% mediante regresión lineal

Term	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	0,241715904	0,009124722	26,49022112	2,03408E-43
ligaESP	0,007572311	0,012517896	0,604918774	0,546807539
ligaFRA	0,041316163	0,014013597	2,94829109	0,004102243
ligaGER	0,007035205	0,01222719	0,575373838	0,566523682
ligaITA	0,023479829	0,011577031	2,028138972	0,045604964
temporada_centrado	-0,002155405	0,001224594	-1,760097645	0,081905988
Valor_Media_centrado	4,85168E-11	8,69136E-11	0,558219285	0,578128429
Valor_DesvTip_centrado	-2,39963E-11	1,22432E-10	-0,195997679	0,84506906
	R.squared	**Adj.R.squared**	**F.statistic**	**p.value**
	0,2779	0,2198	4,7834	0,000137976

Fuente: Elaboración propia con R

Tabla 8. Resultado del modelo de Machine Learning aplicando logaritmos para el índice Draw% mediante regresión lineal

Term	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	10,02282632	11,84740331	0,845993511	0,399877245
ligaESP	0,062743239	0,051766626	1,212040353	0,228778351
ligaFRA	0,160095167	0,075982009	2,107014132	0,037995153
ligaGER	0,029006784	0,057963781	0,500429466	0,618036289
ligaITA	0,090127295	0,048945504	1,841380452	0,068973714
temporada	-0,005756137	0,0066023	-0,871838194	0,385696313
log(Valor_Media)	0,118178058	0,098180677	1,203679402	0,231978444
log(Valor_DesvTip)	-0,112227999	0,065444483	-1,714858072	0,089931319
	R.squared	**Adj.R.squared**	**F.statistic**	**p.value**
	0,2842	0,2266	4,9351	9,91817E-05

Fuente: Elaboración propia con R

En la Tabla 9 se encuentra una salida de R Studio con los resultados del modelo de *Machine Learning para* el índice RSD como variable dependiente y como explicativas las variables que miden el desequilibrio competitivo y las económicas. En el modelo de la Tabla 9 se ha procedido a centrar las variables económicas y a restar la primera temporada; estas transformaciones permiten una mejor estimación e interpretación de los coeficientes del modelo.

Todas las estimaciones están comparadas en relación con el RSD de la Premier League.

A continuación, se procede a interpretar el resultado:

En primer lugar, se observa que el intercepto corresponde a la Premier League (**ligaENG**), tiene un valor de 204.2 que se interpreta como el índice de la Premier en la primera temporada cuando el valor medio coincide con su media y la desviación típica también. Dado que el p-valor tiene un resultado de prácticamente cero, es estadísticamente significativo, y por tanto se debería de tener en cuenta como un indicador preciso del balance competitivo para la primera temporada y cuando las variables están centradas.

Todas las estimaciones están comparadas en relación con el RSD de la Premier League.

En cuanto a las estimaciones de los índices elaborados para cada liga:

- En primer lugar, en La Liga (**ligaESP**)**,** tiene un coeficiente negativo de -4.231 y un p-valor de 0.652166>0.05 no siendo estadísticamente significativa la diferencia respecto la Premier, lo que significa que no existe suficiente evidencia para determinar que el desequilibrio competitivo en la liga española medido por el RSD es diferente en comparación con la liga inglesa.

- En cuanto a la Ligue 1 (**ligaFRA**), presenta un coeficiente de -44.06 y un p-valor de 0.0000626<0.05, lo que sugiere que la liga francesa tiene un RSD más bajo que la inglesa, y por tanto a mayor equilibrio competitivo respecto a la liga inglesa.
- En la Bundesliga (**ligaGER),** el coeficiente es de -29.50 y un p-valor de 0.001756<0.05 lo que implica que es estadísticamente significativa la diferencia respecto de la Premier y por tanto implica que esta liga tiene un RSD más bajo que la inglesa y en consecuencia mayor equilibrio competitivo.
- Por último, la Serie A (**ligaITA),** con un coeficiente de -5.809 pero con un p-valor de 0.503746>0.05 indica que este coeficiente no es estadísticamente significativo y no se puede saber si existe diferente RSD en comparación con la liga inglesa y por tanto diferencias con el Balance Competitivo.
- Por tanto, la liga francesa y alemana tienen menor RSD que la inglesa y por ello mayor balance competitivo, mientras que la española e italiana son similares a la inglesa.

Por otro lado, el coeficiente de **temporada** 3.215 siendo este positivo, y con un p-valor de 0.000704<0.05 estadísticamente significativo indicando que el paso del tiempo implica una tendencia hacia un mayor RSD y por tanto menor balance competitivo a lo largo de las temporadas.

En cuanto a las variables económicas, el valor de mercado medio durante las temporadas analizadas (**Valor_Media)** tiene coeficiente de -1.399e-07 y la desviación típica de este (**Valor_DesvTip)** tiene coeficiente de 5.914e-08, siendo solo el **Valor_Media** es estadísticamente significativa con un p-valor de 0.034022<0.05 y por tanto el valor de mercado medio de los equipos muestra una mejora del RSD y por tanto del Balance Competitivo. Mientras que **Valor_DesvTip,** no afecta.

En cuanto al ajuste del modelo, los valores de R^2 son de 0.4433 y el ajustado de 0.3985, lo que sugiere que el modelo tiene un ajuste moderado hay una variabilidad del 60.15% que el modelo no explica completamente, indicando que hay otros efectos de factores no explicativos que no captura el modelo.

Finalmente se observa que tanto el estadístico F como el p-valor tiene relevancia estadística y por tanto al menos una de las variables explicativas está relacionada con el RSD y por tanto el modelo es útil.

En el caso de la Tabla 10, se han aplicado logaritmos al RSD y a las variables económicas para conocer el incremento porcentual del índice por cada aumento porcentual de estas variables. En primer lugar, se observa que, de las ligas estadísticamente significativas, solo lo son la francesa y alemana, con coeficiente negativo por ello existe un mayor equilibrio competitivo y mayor RSD que en la Premier.

En cuanto a las variables económicas, solo la desviación típica es estadísticamente significativa, esta muestra que por cada 1% que crece esta, hay un aumento del RSD de 14.7195%, lo que implica un empeoramiento del Balance Competitivo.

Tabla 9. Resultado del modelo de Machine Learning para el índice RSD mediante regresión lineal

Term	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	204,1557058	6,818794404	29,94014684	1,36666E-47
ligaESP	-4,231190863	9,354472522	-0,452317419	0,652166148
ligaFRA	-44,06063702	10,47219139	-4,20739417	6,25939E-05
ligaGER	-29,49962904	9,13723119	-3,228508553	0,001755806
ligaITA	-5,808510685	8,651375547	-0,671397358	0,503746245
temporada_centrado	3,21540776	0,915124355	3,513629316	0,000703806
Valor_Media_centrado	-1,39887E-07	6,49495E-08	-2,153789262	0,03402236
Valor_DesvTip_centrado	5,91412E-08	9,14917E-08	0,646411254	0,51971468
	R.squared	**Adj.R.squared**	**F.statistic**	**p.value**
	0,4433	0,3985	9,8980	4,99531E-09

Fuente: Elaboración propia con R

Tabla 10. Resultado del modelo de Machine Learning aplicando logaritmos para el índice RSD mediante regresión lineal

Term	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	-17,69591521	12,67437879	-1,396195861	0,166207926
ligaESP	-0,021322053	0,055380053	-0,385013226	0,701167266
ligaFRA	-0,198674385	0,081285725	-2,444148535	0,016537829
ligaGER	-0,116397256	0,062009783	-1,8770789	0,063858741
ligaITA	0,004647261	0,052362011	0,08875253	0,929482571
temporada	0,011676492	0,007063155	1,653155184	0,1019036
log(Valor_Media)	-0,173616304	0,10503391	-1,652954791	0,10194449
log(Valor_DesvTip)	0,147194632	0,070012655	2,102400365	0,038407949
	R.squared	**Adj.R.squared**	**F.statistic**	**p.value**
	0,4481	0,4037	10,0923	3,51536E-09

Fuente: Elaboración propia con R

4.2.2. Análisis del método Random Forest

La Tabla 11, junto a la Figura 6, muestra una salida de R en la cual se observa el resultado dado para el SHICB haciendo uso del algoritmo Random Forest. En este caso SHICB es la variable dependiente y las demás variables son explicativas: liga, temporada y variables económicas.

En primer lugar, el error cuadrático medio de los residuos es de 2.24077, lo que indica la diferencia promedio al cuadrado entre los valores observados y predichos por el modelo. En cuanto a la varianza explicada por este, se sitúa en el 31.55% lo cual es un ajuste

moderado, esto se puede interpretar como que aproximadamente un tercio de la variabilidad en el SHICB se explica por las variables incluidas en el modelo, aunque hay que tener en cuenta que un 68.55% de la variabilidad no está siendo capturada por el modelo, por lo que pueden existir otros factores importantes que afectan al SHICB.

Las variables más importantes en este modelo (Tabla 11) son las económicas, **Valor_Media_centrado** y **Valor_DesvTip_centrado,** esto puede deberse a que la distribución del talento es desigual en los equipos y por tanto afecta significativamente a la competitividad. La variable **temporada** también hay que tenerla en cuenta, pues puede ser que el paso del tiempo sea un factor que afecte negativamente a la competitividad también, en menor medida afectan las particularidades de cada **liga** al SHICB.

Observando las Tablas 12 y 13, junto a las Figuras 7 y 8, los resultados obtenidos en todos los modelos de RandomForest para los demás índices, ofrecen el mismo resultado que el resultado del SHICB descrito anteriormente, respecto a la importancia de las variables económicas y viceversa. En el caso del índice Draw%, este ofrece un error cuadrático medio de 0.0009, junto a una varianza explicada del modelo de 7.02%, en este caso muy limitada y por tanto este modelo no se tendrá en cuenta. Finalmente, el RSD este tiene un error cuadrático medio de 507.7273, el mayor entre los índices, además la varianza explicada para este modelo de 30.87%.

Tabla 11. Resultado del modelo de Machine Learning para el índice SHICB mediante Random Forest

Variable	**IncNodePurity**
liga	41,99804641
temporada_centrado	55,25341361
Valor_Media_centrado	76,62765284
Valor_DesvTip_centrado	85,06409193
	Mean of squared residuals: 2.24077
	% Var explained: 31.55

Fuente: Elaboración propia con R

Tabla 12. Resultado del modelo de Machine Learning para el índice Draw% mediante Random Forest

Variable	**IncNodePurity**
liga	0,010502561
temporada_centrado	0,016557996
Valor_Media_centrado	0,025648253
Valor_DesvTip_centrado	0,023690966
	Mean of squared residuals: 0.0009427783
	% Var explained: 7.02

Fuente: Elaboración propia con R

Tabla 13. Resultado del modelo de Machine Learning para el índice RSD mediante Random Forest

Variable	**IncNodePurity**
liga	9152,765862
temporada_centrado	12379,69533
Valor_Media_centrado	17156,0234
Valor_DesvTip_centrado	18896,45915
	Mean of squared residuals: 507.7273
	% Var explained: 30.87

Fuente: Elaboración propia con R

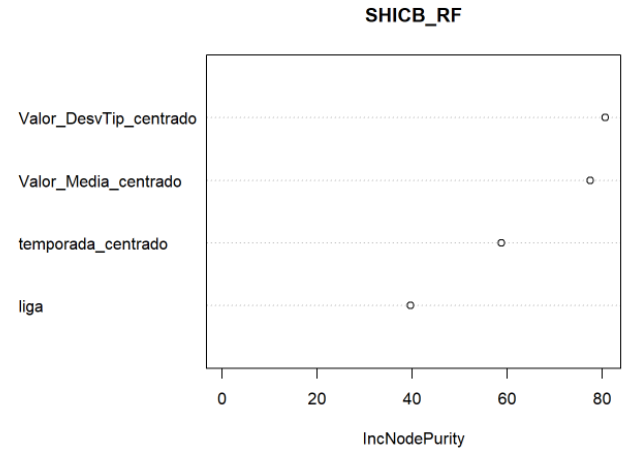

Figura 6. Resultado del modelo de Machine Learning para el índice SHICB mediante Random Forest

Fuente: Elaboración propia con R

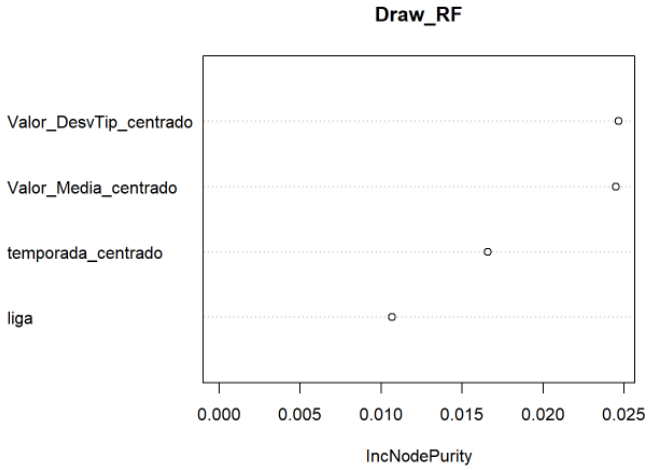

Figura 7. Resultado del modelo de Machine Learning para el índice Draw% mediante Random Forest

Fuente: Elaboración propia con R

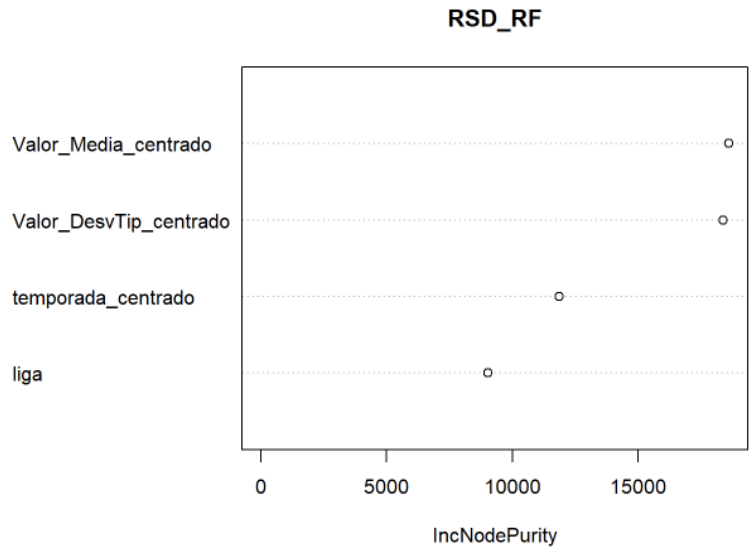

Figura 8. Resultado del modelo de Machine Learning para el índice RSD mediante Random Forest

Fuente: Elaboración propia con R

5. Conclusiones

El objetivo principal de este TFG fue conocer la liga más desequilibrada y por ende a largo plazo la posiblemente menos rentable entre las cinco grandes ligas.

Para alcanzarlo, se establecieron una serie de objetivos secundarios, los cuales se han cumplido cada uno de ellos. El primero de ellos, se concluye con la obtención de los datos de las cinco grandes ligas europeas para las temporadas 1999/2000 a la 2022/2023 de manera satisfactoria gracias al paquete de R WorldFootballR y al código de web scraping utilizado para obtener los datos a partir de la web de Fbref.

Para alcanzar el segundo objetivo, se han elaborado los distintos índices para medir el grado de desequilibrio competitivo en cada liga y temporada, donde a priori se muestra que existe desequilibrio competitivo en las cinco grandes ligas y su tendencia creciente. Además, se obtiene el gráfico de dominancia el cual muestra que la liga más competitiva es la Ligue 1 (liga francesa) y la menos competitiva la Bundesliga (liga alemana).

Respecto al tercer objetivo secundario, se han obtenido la media y desviación típica para cada liga y temporada gracias a la web de Transfermarkt.

Por último, se ha logrado elaborar un modelo de *Machine Learning* mediante los índices calculados anteriormente y las variables económicas, se han aplicado los métodos de regresión lineal y el algoritmo de Random Forest. En regresión lineal se ha usado la Premier League como punto de referencia; tal y como muestra dicha regresión las ligas en general muestran balances competitivos significativamente diferentes en todos los índices analizados.

En cuanto al impacto de variables explicativas del desequilibrio competitivo medido con los índices mediante regresión lineal, se observa respecto a la de variables económicas que sí existe efecto significativo en las ligas estudiadas. Las ligas con coeficiente negativo en SHICB tienen menor desequilibrio competitivo en comparación con la Premier. Las variables económicas, aunque son estadísticamente significativas en algunos casos, tienen un impacto influenciado por las unidades, por ello se ha aplicado logaritmos a estas y al índice analizado en cada caso para conocer el aumento porcentual que sufre el índice por cada aumento porcentual de estas variables. Pese a esto, aunque el valor de mercado medio y desviación típica influye en el equilibrio competitivo, existen algunos factores no capturados en el modelo.

En cuanto a la comparación entre ligas en regresión, como se ha dicho la Premier League, se usa como referencia por las dummies creadas en el método de regresión lineal. En comparación, La Liga, Ligue 1 y Bundesliga, muestran diferentes grados de mayor equilibrio competitivo en la mayoría de los índices con relación a la Premier League, con la Ligue 1 mostrando tendencia a un mayor equilibrio competitivo. En cuanto a la Serie A, no muestra relación significativa en la mayoría de los índices, lo que indica que no hay suficiente evidencia de que el desequilibrio competitivo en la liga sea diferente de la inglesa.

En cuanto al paso del tiempo, este no muestra una tendencia clara hacia un mayor o menor balance competitivo en las ligas, excepto en el índice RSD, donde se observa una

tendencia hacia un menor RSD y por tanto mayor equilibrio competitivo a lo largo de las temporadas.

En cuanto a los modelos del algoritmo de Random Forest, destaca la importancia de las variables económicas en todos los índices los cuales muestran un ajuste moderado, excepto en el Draw%, el cual es limitado y no es adecuado para explicar la variabilidad de los empates. La siguiente variable más importante es temporada, lo que sugiere que el paso del tiempo puede influir en la rivalidad en las ligas.

En este modelo, la influencia de las variables económicas sugiere que la competitividad de las ligas puede estar afectada por la distribución desigual del talento y los recursos económicos de los equipos, en este caso las particularidades de cada liga tienen un impacto menor en la competitividad.

De análisis conjunto de los resultados de los métodos de regresión lineal y del Random Forest, se puede concluir que las variables económicas son significativas para ambos modelos, lo que indica que la distribución de talento y recursos económicos tienen un impacto en la competitividad de las ligas. Ambos métodos muestran un ajuste moderado, lo que significa que son capaces de explicar una parte de la variabilidad en los índices, pero también dejan una proporción significativa de la variabilidad por explicar. La variable temporada aparece como relevante en ambos métodos, por lo que sugiere que el paso del tiempo puede influir en la competitividad de las ligas, aunque el impacto varía entre los índices.

En cuanto a las limitaciones encontradas, una gran proporción de la variabilidad en los índices no está siendo captada por ninguno de los modelos, lo que indica que hay otros factores importantes que afectan a la rivalidad de las ligas que no está incluido en los modelos actuales. Siendo que el modelo de Draw% queda descartado por este mismo problema.

Finalmente, remarcar que los objetivos se han cumplido. Se conoce que la liga más desbalanceada para las temporadas analizadas es la Premier League. Para las implicaciones para futuras investigaciones, estos modelos podrían beneficiarse de la inclusión de más variables o la aplicación de otros métodos de modelado para capturar mejor la complejidad del equilibrio competitivo en las ligas de fútbol. También podría estudiarse si el aumento de empates es fruto de un aumento de la competitividad y si esta competitividad es atractiva para el espectador. Además, todo este análisis realizado es fácilmente reproducible gracias a R y su interfaz RStudio, software de código abierto, elaborado totalmente con datos reales obtenidos por Fbref y Transfermarkt y que abre un análisis novedoso e interesante pudiendo además aplicarse a otros muchos deportes.

Bibliografía

Bankinter. (13 de 02 de 2024). *UEFA Champions League 2023/24: ¿Cuánto dinero gana cada equipo?* Obtenido de https://www.bankinter.com/blog/empresas/champions-league-dinero-ganan-equipos-victoria-empate-eliminatorias-campeon#:~:text=%2D%20Ser%20campe%C3%B3n%3A%20El%20equipo%20que,5%20millones%20de%20euros%20adicionales.

Budzinski, O. &. (2014). The behavioural economics of competitive balance: Implications for league policy and championship management.

Buzzacchi, L. S. (2003). Equality of Opportunity and Equality of Outcome: Open Leagues, Closed Leagues and Competitive Balance. *Journal of industry, competition and trade, 62*(3), 167-186.

El País . (2024). Obtenido de https://elpais.com/deportes/futbol/2023-12-04/la-premier-eleva-sus-ingresos-por-derechos-televisivos-en-un-4.html

El-Hodiri, M. &. (1971). An economic model of a professional sports league. *Journal of Political Economy, 79*(6), 1302-1319.

Fernández, P. S. (2019). "Finanzas del deporte: Fuentes de ingreso y regulación financiera en el fútbol europeo". *academia.edu*.

Fort, R. M. (2016). Uncertainty by regulation: Rottenberg′ s invariance principle. *Research in Economics, 70*(3), 454-467.

Goossens, K. (2006). Competitive balance in European football: Comparison by adapting measures: National measure of seasonal imbalance and top 3. *Rivista di Diritto ed Economia, 2*(2), 77-122.

Groot, L. (2008). Economics, uncertainty and European football: Trends in competitive balance. . *Edward Elgar Publishing*.

Haugen, K. (2008). Point score systems and competitive imbalance in professional soccer. . *Journal of Sports Economics, 9*(2), 191-210.

Kringstad, M. &. (2018). Is gender a competitive balance driver? Evidence from Scandinavian football. *Cogent Social Sciences, 4*(1), 1439264-1439215.

La Liga. (2024). *Reparto de los Derechos de TV*. doi:10.1080/13504851.2021.2023088

Liaw, A. &. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22. Obtenido de https://CRAN.R-project.org/doc/Rnews/

Mitchie, J. &. (2004). Competitive balance in football: Trends and effects (Research paper 2004 No. 2). *University of London, Football Governance Research Centre*.

Mondal, S. (2023). She kicks: The state of competitive balance in the top five women's football leagues in Europe. *Journal of Global Sport Management, 8*(1), 432-454.

Neale, W. (1964). The peculiar economics of professional sports. A contribution to the theory. *The Quarterly Journal of, 1*(1-14).

Noll, R. G. (1974). Government and the sports business: Papers prepared for a conference of experts, with an introduction and summary. *Brookings Institution*.

Plumley, D. R. (2018). Mind the gap: an analysis of competitive balance in the English football league system. *International Journal of Sport Management and Marketing, 18(5)*, 357-375.

R Core Team. (2023). Computing, R: A Language and Environment for Statistical. Obtenido de https://www.R-project.org/

Ramchandani, G. P. (2018). A longitudinal and comparative analysis of competitive balance in five European football leagues. *Team Performance Management:An International Journal, 24*(5-6), 265-282.

Ramchandani, G. P. (2019). Does size matter? An investigation of competitive balance in the English Premier League under different league sizes. *Team Performance Management: An International Journal, 25*(3/4), 162-175.

Rocke, K. (2019). Competitive balance within CONCACAF: a longitudinal and comparative descriptive review of the seasons 2002/2003–2017/2018. *Managing Sport and Leisure, 24*(6), 445-460.

Rosen, S. (1981). The economics of superstars. *The American economic review, 71*(5), 845-858.

Rottenberg, S. (1956). The baseball players' labor market. *Journal of political economy, 64*(3), 242-258.

RStudio Team. (2023). *RStudio: Integrated Development for R. RStudio.* PBC, Boston. Obtenido de http://www.rstudio.com/.

Silva, C. A. (2018). Competitive balance in football: A comparative study between Brazil and the main European leagues. *Journal of Physical Education, 29*(1), 1-11.

Stephen Dobson, J. G. (2011). *The Economics of Football.* Cambridge University Press.

Szymanski, S. (2001). Income inequality, competitive balance and the attractiveness of team sports: Some evidence and a natural experiment from English soccer. *The Economic Journal, 111*(469), 69-84.

Universidad Europea. (2023). *What is Financial Fair Play?* Obtenido de https://universidadeuropea.com/en/blog/what-is-financial-fair-play/

Vrooman, J. (2015). Economics of Sport. *Scottish Journal of Political Economy, 62*(1), 90-115.

Walvin, J. (2014). *The people's game: the history of football revisited.* Random House.

Weisberg, J. F. (2019). An {R} Companion to Applied Regression. (Sage, Ed.) Obtenido de https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Zivkovic, J. (2024). *worldfootballR: Extract and Clean World Football (Soccer) Data* (R package version 0.6.5.0003 ed.). Obtenido de https://github.com/JaseZiv/worldfootballR

Anexo I. Objetivos de Desarrollo Sostenible

Tabla 1. Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible

Objetivos de Desarrollo Sostenibles	**Alto**	**Medio**	**Bajo**	**No Procede**
ODS 1. **Fin de la pobreza**				**X**
ODS 2. **Hambre cero**				**X**
ODS 3. **Salud y bienestar**				**X**
ODS 4. **Educación de calidad**				**X**
ODS 5. **Igualdad de género**				**X**
ODS 6. **Agua limpia y saneamiento**				**X**
ODS 7. **Energía asequible y no contaminante**				**X**
ODS 8. **Trabajo decente y crecimiento económico**		**X**		
ODS 9. **Industria, innovación e infraestructuras**	**X**			
ODS 10. **Reducción de las desigualdades**				**X**
ODS 11. **Ciudades y comunidades sostenibles**				**X**
ODS 12. **Producción y consumo responsables**				**X**
ODS 13. **Acción por el clima**				**X**
ODS 14. **Vida submarina**				**X**
ODS 15. **Vida de ecosistemas terrestres**				**X**
ODS 16. **Paz, justicia e instituciones sólidas**				**X**
ODS 17. **Alianzas para lograr objetivos**				**X**

Fuente: Elaboración propia.

Este TFG se relaciona estrechamente con dos de los Objetivos de Desarrollo Sostenible (ODS), especialmente enfocados en el ámbito del deporte y la industria del fútbol.

El principal vínculo se establece con el Objetivo 9: Industria, innovación e infraestructuras. El desarrollo y aplicación de modelos de *Machine Learning* para analizar el balance competitivo en las ligas de fútbol europeas representan un avance significativo en términos de innovación tecnológica en el ámbito deportivo. Esta iniciativa contribuye específicamente a la meta 9.5 de la Agenda 2030, que busca aumentar la investigación científica y fortalecer la capacidad tecnológica en diversos sectores industriales. Al aplicar estas técnicas analíticas avanzadas, se mejora la comprensión de los factores que influyen en la competitividad en el fútbol, lo que permite

tomar decisiones más informadas para promover la innovación y el crecimiento en esta industria.

Además, este trabajo también se vincula con el Objetivo 8: Trabajo decente y crecimiento económico. La investigación sobre el balance competitivo en las ligas de fútbol contribuye al desarrollo económico sostenible en el ámbito deportivo. Al entender mejor las dinámicas de competencia en el fútbol, se pueden identificar oportunidades para mejorar la gestión de los recursos y promover la eficiencia en las organizaciones deportivas. Esto favorece la creación de empleo y el crecimiento económico en el sector del deporte, alineándose con la meta 8.2 de la Agenda 2030, que busca aumentar la productividad económica a través de la innovación y la modernización tecnológica.

En conclusión, este TFG destaca por su contribución a la innovación tecnológica y al desarrollo económico sostenible en el contexto de la industria del fútbol. Al aplicar modelos analíticos avanzados para entender mejor el balance competitivo en las ligas europeas, se promueve la toma de decisiones informadas que pueden impulsar la eficiencia y el crecimiento en el deporte. Si bien podría haber algunas conexiones menores con otros ODS, la relación más significativa se establece con los Objetivos 9 y 8, enfocados en la industria, la innovación y el crecimiento económico en el ámbito deportivo.