

Diversificación en la configuración de los equipos de la Primera División española de fútbol

MANUEL A. ESPITIA-ESCUER

Departamento de Dirección y Organización de Empresas, UNIVERSIDAD DE ZARAGOZA, ESPAÑA. E-mail: espitia@unizar.es

LUCÍA ISABEL GARCÍA-CEBRIÁN

Departamento de Dirección y Organización de Empresas, UNIVERSIDAD DE ZARAGOZA, ESPAÑA. E-mail: lgarcia@unizar.es

RESUMEN

El diseño y configuración de los equipos deportivos se realiza de acuerdo con los objetivos establecidos por la institución a la que pertenecen. En este trabajo se analiza la configuración de los equipos y la eficacia en la consecución de éxitos deportivos tomando como referencia el índice de diversificación de Gollop y Monahan.

Se plantea un modelo en el que los componentes de dicho índice se incorporan como variables explicativas de los resultados obtenidos por los equipos. La estimación del modelo se realiza con la técnica de clasificación 'regression tree', la muestra queda configurada por todos los equipos de las ediciones de la Liga de Primera División, de 2003/04 a 2010/11. Los resultados obtenidos en este trabajo, ponen de manifiesto especialmente la relevancia del equipo y las estrategias como modelo dominante en los equipos que han alcanzado el máximo nivel en esta competición.

Palabras clave: Diversificación en equipos deportivos, índice de diversificación, Liga Española de Fútbol.

Diversification in the Team Configuration of the Spanish Football First Division

ABSTRACT

The design and configuration of sports teams is made according to the goals set by the institution to which they belong. This paper analyzes the team configuration and the effectiveness in the pursuit of sporting successes based on the Gollop and Monahan diversification index.

In the model, the components of the index are incorporated as explanatory variables of results obtained. The estimation of the model is made with the technique of classification 'regression tree', the sample is configured for all teams in the editions of the First Division League of 2003/04 to 2010/11. Results from this work highlights especially the relevance of the teams strategies as a dominant model in teams that have reached the competition highest level.

Keywords: Team Sport Diversification, Diversification Index, Football Teams, Spanish Liga.

Clasificación JEL: L83, L25, J24

Artículo recibido en marzo de 2012 y aceptado en junio de 2012

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref. e-30209

1. INTRODUCCIÓN

Los equipos deportivos se configuran de acuerdo con unas determinadas preferencias definidas por sus directivos o directores técnicos con el fin de alcanzar las metas deportivas establecidas por la institución. Públicamente las metas se difunden de forma retórica en torno a objetivos tales como llegar a participar en competiciones europeas, mantener el equipo en puestos sin problemas de cara al final de la temporada, etc.

El espectáculo que genera su actuación en el terreno de juego resulta de combinar las aptitudes de sus componentes de distintas formas y ofrecer un juego de características diferentes al de sus rivales. Estas diferencias las percibe y valora el público, quien, incluso en conversaciones informales, comenta las distintas estrategias de juego adoptadas a nivel competitivo, Esteban y Coll (2003), por cada equipo. El espectáculo ofrecido se percibe por parte del público como un espectáculo de calidad, en la medida en la que se pongan en práctica estrategias que den como resultado no solamente éxitos deportivos, sino también jugadas de mérito, combinaciones y movimientos entre jugadores que encierren cierta belleza plástica en su ejecución.

El estudio de la actividad económica de los clubs deportivos tiene ya una larga tradición, puesto que se trata de organizaciones que ofrecen un servicio (espectáculo) a un mercado (los aficionados) por el que reciben unos ingresos y para el cual utilizan unos recursos productivos tales como instalaciones de entrenamiento y competición, gimnasios y otras instalaciones específicas, pero el elemento fundamental es la 'mano de obra especializada' (jugadores y técnicos) con la que se configura el equipo con el que se participa en las diversas competiciones deportivas.

Bajo una óptica de gestión del equipo, la configuración del mismo según el tipo de jugadores incorporados permite una serie de variantes en el juego que quedan limitadas por la plantilla y que obedecen a los diseños estratégicos preconcebidos por los técnicos. Aquellos equipos con mayor número de jugadores podrán 'a priori' poner en práctica más alternativas en el juego que las plantillas con menor número de efectivos disponible. Sin embargo, no es tanto el número de jugadores disponible como la versatilidad de los mismos lo que permite a los técnicos diseñar variantes en el planteamiento de los distintos partidos de la competición. La caracterización de cada jugador permitiría un enfoque diferencial de los equipos en función de las características de sus miembros, sin embargo, el concepto y la tecnología de equipo, suma de individualidades con un objetivo de grupo, son el soporte de los clubs deportivos.

La medición de los resultados, la rentabilidad o la evaluación de la gestión llevada a cabo no son las únicas áreas en que se pueden aplicar las herramientas propias de la Economía. Aunque resultaría menos evidente, la propia actuación deportiva en el terreno de juego también es susceptible de valoración propia de

la Economía. Así, podemos encontrar trabajos que intentan predecir los resultados deportivos como los de Klaassen y Magnus (2003), Koning *et al.* (2003) y Dobson y Goddard (2003). Asimismo, hay una corriente que se dedica a evaluar los resultados deportivos obtenidos, como puede apreciarse en los trabajos de Lewis (2005), Carmichael y Thomas (2005), Caliendo y Radic (2006), Gracia y Rodriguez (2002), Espitia y Garcia (2004) y (2006). También se encuentran trabajos que tratan sobre la asignación de los jugadores a las posiciones en el terreno de juego, tales como Boon y Sierksma (2003) y Moschini (2003).

En este trabajo se toma como muestra los equipos de fútbol que han participado en la Liga de Primera División en las temporadas que van de la 2003-2004 a la 2010-2011. Se va a calcular el valor de un índice de diversificación que mida el tipo de juego adoptado y se va a relacionar con los resultados deportivos obtenidos con el fin de detectar configuraciones de equipo que se puedan identificar como de éxito en la competición.

El marco de análisis que se propone en este trabajo se basa en el uso del índice de Gollop y Monahan (1991) (en adelante GM) que proporciona tanto un valor agregado que mide la diversificación para cada uno de los equipos como la descomposición en sus componentes. En este sentido se va a tomar la propuesta de Gollop y Monahan calculando un índice de diversificación formado por tres componentes que tienen en cuenta, respectivamente, el número de productos, la igualdad en su distribución en la actividad de la organización y el grado de similitud entre esos productos; puesto que el trabajo de Gollop y Monahan (1991) se centra en la industria manufacturera, se hace necesaria una adaptación al caso de los equipos de fútbol. A partir de esos valores se pretende detectar los modelos competitivos diseñados por los técnicos de los equipos que han participado en las últimas ediciones de la Liga de Primera División y su relación con los resultados alcanzados en esa competición, es decir, va a ser utilizado como una variable explicativa del nivel de eficacia alcanzado por los equipos que han participado en dicha competición.

La estructura del trabajo es la siguiente: en el segundo apartado se va a mostrar la solución adoptada al problema del índice a utilizar; en el tercer apartado se presenta la metodología y la base de datos utilizadas; finaliza el trabajo con la presentación de los resultados obtenidos y las conclusiones.

2. DIVERSIFICACIÓN EN LA CONFIGURACIÓN DE EQUIPOS DE FÚTBOL

Las estrategias de diversificación han sido analizadas desde la óptica empresarial con notable intensidad en los últimos años. La diversificación empresarial, entendida en términos generales como la diversidad en la oferta de productos, ha sido objeto de estudio fundamentalmente desde dos líneas diferenciadas. La primera línea ha tomado como base la necesidad de una definición

correcta del término diversificación y de una medida que cumpliera los requisitos necesarios para ofrecer información fiable y comparable entre empresas. A esta línea pertenecerían los trabajos de Hoskisson et al (1993), Troutt y Acar (2005), Gollop y Monahan (1991) y Pehrsson (2006). Los trabajos pertenecientes a la segunda línea de investigación analizan la influencia que las decisiones de diversificación han tenido en los resultados de la empresa; como ejemplos, se podrían citar los trabajos de Graham, Lemmon y Wolf (2002), Jacquemine y Berry (1971), Markides y Willimanson (1994) y Rajan, Servaes y Zingales (2000). No obstante a pesar de haberse adoptado diversas ópticas, las conclusiones no resultan generalizables en cuanto a recomendaciones estratégicas.

En la literatura se han propuesto índices muy variados que permiten aproximar a través de una medida única las características de las estrategias de diversificación desarrolladas por las empresas. A partir de los índices sencillos de concentración, han ido apareciendo índices más complejos como el índice Herfindahl o los índices de Berry (1971) y de Gort (1962) entre otros. Gollop y Monahan (1991) proponen un índice de diversificación más completo y capaz de incluir a los anteriores como casos particulares y cumple las tres propiedades fundamentales para ser considerado como un índice válido:

1. Varía directamente con el número de bienes diferentes producidos por la organización.
2. Varía inversamente con la desigualdad en la distribución de los productos en la composición de la oferta total.
3. Varía directamente con la heterogeneidad de los productos.

En su aplicación al análisis de las plantas productivas, empresas o industrias, el índice GM queda especificado como:

$$GM = \frac{1}{2} ((1-1/n)) + \frac{1}{2} \sum_i (1/n^2 - s_i^2) + \frac{1}{2} \sum_i \sum_k s_i s_k \sigma_{ik}$$

donde n es el número de productos distintos que produce la empresa, s_i es la proporción que representa cada producto i en las ventas totales y σ_{ik} es una variable representativa de la no similitud de los productos que se calcula como

$$\sigma_{ik} = \left(\sum_j \frac{|\omega_{kj} - \omega_{ij}|}{2} \right)^{1/2}$$

donde, ω_{kj} es la proporción que el input j representa en el coste total de los factores en el bien k .

De la expresión anterior se observa que el índice puede quedar descompuesto en tres componentes que tienen significado por sí mismos y que confieren al índice un potencial de uso de notable interés en el ámbito académico:

$$GM = N + S + H$$

Donde:

$$N = \frac{1}{2} (1 - 1/n); \quad S = \frac{1}{2} \sum_i (1/n^2 - s_i^2); \quad H = \frac{1}{2} \sum_i \sum_k s_i s_k \sigma_{ik}$$

El índice GM toma valores que van desde 0 a 1, la diversificación perfecta se corresponde con el valor 1 y la máxima concentración con el valor 0. Los tres componentes N, S, y H crecen con el número de productos, con la homogeneidad en la distribución y con la diferencia entre los productos.

El primer componente del índice (N) se identifica como el componente número de productos de la empresa y toma valores que oscilan en el intervalo $[0 \text{ y } \frac{1}{2}]$, de forma que cuanto mayor sea el valor de n, el número de unidades de negocio, mas próximo a $\frac{1}{2}$ estará el valor del componente N del índice.

El segundo componente del índice (S) refleja la distribución de los negocios de la empresa entre sus áreas de actividad. Como se puede apreciar, resulta de la diferencia entre dos índices de Herfindahl: el que se obtendría en la empresa si la distribución de la actividad entre los negocios fuera homogénea ($1/n^2$), y el índice de Herfindahl ($\sum_i s_i^2$) de la distribución real de la actividad de la empresa entre sus negocios. Este componente toma valores negativos dentro del intervalo $[-\frac{1}{2} (1 - 1/n) < S \leq 0]$, de forma que cuanto menor sea el valor de S mayor será la asimetría entre los negocios de la empresa.

El tercer componente del índice (H) identifica la heterogeneidad de las unidades de negocio, siendo mayor su valor cuanto mayor sea la heterogeneidad entre los negocios de la empresa. Los valores de este componente se sitúan dentro del intervalo que queda definido por los valores positivo y negativo del complemento a la unidad de los índices de Herfindahl reales obtenidos anteriormente y podría interpretarse como el que resulta de la distribución de los negocios sin heterogeneidad entre ellos,

$$[-\frac{1}{2} (1 - \sum_i s_i^2) \leq H \leq \frac{1}{2} (1 - \sum_i s_i^2)]$$

La diversificación como estrategia en la configuración de los equipos deportivos puede entenderse bajo una doble óptica, por una parte la que resulta de diseñar una plantilla con una gama de jugadores con características distintas con los que poder desarrollar planteamientos diferentes, o bien, por otra parte la que resulta de configurar la plantilla con jugadores polivalentes que sean capaces de ejecutar los diversos planteamientos estratégicos diseñados por los técnicos del equipo. En el primer modelo de diseño prevalecen el jugador y sus características, mientras que en el segundo prevalece la estrategia de equipo diseñada por los técnicos.

Para cumplir con los objetivos de este trabajo, se han adaptado las variables que intervienen en el índice de diversificación de GM al caso de la competición analizada. Así, en el término relativo al número de componentes, \underline{n} se toma como el número de jugadores de cada equipo que han participado en la compe-

tición; en el término de la distribución de los componentes, \underline{s} se toma como la cuota de cada jugador en el tiempo de juego del equipo.

A efectos de hacer aplicable esta medida, ω_{kj} , que está definida en términos de coste en la versión original del índice GM, se interpreta en este contexto en términos de esfuerzo, siendo considerado homogéneo para todos los jugadores de la plantilla de cada equipo, de tal modo que se ha calculado como el número de acciones que el jugador k realizar en el área j para cada uno de los equipos de la muestra considerada. El componente σ_{ik} se ha calculado considerando las acciones desarrolladas en cada una de las 17 áreas del campo por cada uno de los \underline{n} jugadores de cada equipo. Cabría interpretarla como la heterogeneidad de los jugadores que conforman la plantilla del equipo, es decir, los jugadores que desarrollan un mayor número de acciones de su equipo contribuyen a la heterogeneidad del juego. Por lo tanto, σ_{ik} representa las diferencias en cuanto a acciones realizadas por un jugador con respecto a los demás del equipo. En consecuencia H , es una medida agregada a nivel de equipo de esa heterogeneidad.

3. METODOLOGÍA Y BASE DE DATOS

La metodología en la que se basa el análisis consiste en la aplicación de la técnica de clasificación CART (Classification and Regresión Tree) propuesta por Breiman, Friedman, Olshen y Stone (1983) y que permite determinar las variables de clasificación relevantes así como los valores que determinan la pertenencia a uno u otro grupo para cada una de las observaciones. Se trata de una técnica de partición binaria recurrente. El término “recurrente” se refiere a que el proceso de partición binaria puede aplicarse una y otra vez. El término “binaria” implica que cada grupo de individuos, representado por un “nodo” en el árbol solo puede dividirse en dos grupos, de forma que cada nodo madre puede generar únicamente dos nodos hijo.

La partición de un nodo del árbol se hace de acuerdo con un criterio que es diseñado para generar nodos hijo que produzcan una suma de cuadrados de errores menor que la del nodo madre. La formación de los nodos hijos a partir de un nodo madre se construye de forma que la disminución de la impureza de todo el conjunto sea máxima. El parámetro CP indica que si la impureza¹ no se reduce en más de un determinado valor de CP, entonces finaliza el desarrollo del árbol.

¹ Por defecto $CP=.01$. Es decir, la reducción en la desviación del nodo terminal debe ser menor del 1% de la desviación inicial. La desviación es una medida del grado de diferencia entre las frecuencias observadas y predichas por el modelo con respecto a la variable dependiente, de forma que a mayor desviación, peor es el modelo. Su cálculo se obtiene como -2 veces el logaritmo neperiano de la verosimilitud del modelo.

Siguiendo a Lewis (2000), la técnica CART es capaz de mostrar complejas interacciones entre las variables que con las técnicas tradicionales quedan ocultas. Este análisis tiene algunas ventajas sobre otros métodos de clasificación, se trata de un método no paramétrico lo que supone que no es necesario hacer supuestos sobre distribución de las variables. Por lo tanto permite manejar muestras de estructura asimétrica, así como series de datos categóricos, sean ordinales o no lo sean.

El modelo al que se propone aplicar la ‘regression tree’ pretende explicar los resultados obtenidos por los equipos en competición en función de los componentes del índice GM, de forma que las variables que hacen referencia a los recursos empleados, la asignación de tareas entre estos recursos (tiempo de juego de cada jugador) y el modo (estrategia o juego individual) en que se realiza esta asignación sean las variables explicativas del resultado obtenido.

La variable dependiente OUTPUT toma los valores de 1 a 4 según el orden de la competición, a los 4 primeros clasificados que son los que alcanzan el nivel que permite participar en la “Champions League” se les asigna el valor 4, a los cuatro siguientes que alcanzan el nivel “Europa League” 3, los 9 siguientes 2 y los 3 equipos que bajan a segunda división 1; las demás variables se corresponden con los componentes del índice GM descritos anteriormente. La utilización de una variable categórica con valores de 1 a 4 para medir el output se justifica por el problema que se genera al utilizar los puntos cuando se toman datos de varias temporadas, el orden de la clasificación también genera problemas al considerar 20 categorías distintas por temporada, de ahí que resulta más conveniente emplear una variable categórica agrupando los equipos por los logros alcanzados.

A través del modelo propuesto se pretende explicar la influencia en la configuración de los equipos y en el uso de los recursos disponibles para alcanzar los objetivos del equipo. La estimación que se propone toma como variable dependiente el Output y como explicativas el índice de GM y sus determinantes. Se han considerado los efectos N, S y H correspondientes a cada uno de los equipos en cada temporada.

La metodología propuesta para llevar a cabo el análisis consiste en la aplicación de la técnica de clasificación ‘regression tree’ que permite determinar las variables de clasificación relevantes así como los valores de tales variables que asignan a los individuos (observaciones) a uno u otro grupo. El análisis se plantea en términos de una serie de datos tratando de forma indiferenciada los valores de las variables en cada una de las temporadas a modo de un pool de datos.

La base de datos utilizada para obtener la información primaria que ha permitido calcular el índice GM y sus efectos es la proporcionada por Geca-Sport en la que se recogen una gran cantidad de variables del desarrollo de los parti-

dos y de la que es posible obtener información detallada por equipos, jugadores y competiciones. Para cada uno de los partidos se incluye información sobre las acciones desarrolladas por cada jugador en cada una de las 17 áreas en que se divide el campo de juego, del tiempo de participación de cada jugador, además de toda una serie de detalles sobre el desarrollo de todos los encuentros de la competición.

4. RESULTADOS

El índice de GM se ha calculado, agregando las acciones de los componentes del equipo por las áreas del terreno de juego. La distribución de los valores puede apreciarse en la Tabla 1 donde se observa que todos los valores están por encima del valor medio del índice. Los valores observados suponen entre el 50 y el 55% del valor máximo del índice, por lo que se podría afirmar que existe una diversificación moderada en los equipos de la Liga.

Tabla 1
Valores de referencia del índice de Gollop y Monahan
para las 8 temporadas analizadas

	GM
Mínimo	0,5069
1er. Q	0,5366
Mediana	0,5386
Media	0,5372
3er. Q	0,5404
Máximo	0,5447

Fuente: Elaboración propia.

Los valores del número de jugadores y de cada uno de los componentes del índice se resumen en la Tabla 2 por medio de los valores de la distribución. Según se aprecia, para el total de las 160 observaciones resultantes de la agregación de las ocho temporadas, los valores de los componentes del índice son muy homogéneos y apenas fluctúan entre los equipos, la excepción la constituye el componente H en el que se observa una gran diferencia entre los valores máximo y mínimo. Sin embargo, destaca la diferencia en el número de jugadores, llegando el valor máximo de la serie a ser casi el doble del mínimo. Lo que muestra una diferencia notable entre los equipos en la configuración de sus plantillas.

Tabla 2
Puntuación y valores de referencia de los componentes del índice GM para todos los equipos de las ligas 03/04 a 10/11

	Puntos	Jugadores	N	S	H
Mínimo	24,00	20,00	0,4750	-0,03530	0,04574
1er. Q	43,00	24,00	0,4792	-0,02977	0,08195
Mediana	49,00	26,00	0,4808	-0,02788	0,08580
Media	52,36	26,20	0,4807	-0,02806	0,08459
3er. Q	60,25	28,00	0,4821	-0,02654	0,08880
Máximo	99,00	35,00	0,4857	-0,02101	0,09582

Fuente: Elaboración propia.

En la Tabla 3 se presentan los valores medios de los componentes del índice GM por temporadas; se aprecia una puntuación media muy similar en cada temporada, al igual que los demás valores de las variables en los que no se aprecian efectos temporales de evolución de las series. Destaca el valor de la variable H en la liga 10/11 que resulta claramente inferior a los demás valores anuales de esta variable, indicando menos heterogeneidad entre las áreas del terreno de juego por término medio en esa temporada.

Tabla 3
Puntuación y valores de los componentes del índice de Gollop-Monahan N, S y H, en promedio por temporada

Liga	Puntos	N	S	H
03/04	52,25	0,481151	-0,028835	0,086435
04/05	52,00	0,481324	-0,029074	0,088491
05/06	51,75	0,481309	-0,028039	0,086385
06/07	52,10	0,480452	-0,027567	0,085281
07/08	52,65	0,481390	-0,026698	0,084195
08/09	52,85	0,480168	-0,027746	0,086791
09/10	52,25	0,478822	-0,028942	0,087715
10/11	53,05	0,480993	-0,027566	0,071441

Fuente: Elaboración propia.

Las estimaciones de los modelos se han realizado con la aplicación de la técnica CART, con el R-Package (rpartOrdinal), siguiendo el método ordinal, adecuado cuando se trata de variables ordinales como la variable utilizada como

medida del Output. El coeficiente CP se ha obtenido en base a la función cuadrática, este método calcula el coste o pérdida asociada con la clasificación de una observación en la clase h o en la clase g como medida de impureza de nodo.

El modelo en el que se toma como variable dependiente el Output del equipo y como variable explicativa en índice GM agregado ha dado como resultado el que se presenta en las Tablas 4, 5 y 6, además de la representación gráfica del Gráfico 1.

La Tabla 4, muestra la configuración de los equipos en cada una de las ramas del árbol y el valor de la variable que determina la pertenencia a una u otra rama. Los estadísticos de la estimación del modelo están recogidos en la Tabla 5. El valor de $CP=0,01$ determina el final de las particiones de las ramas. En la Tabla 6 se muestra la distribución de los equipos por categorías en cada uno de los nodos del modelo. El Gráfico 1 incluye la representación gráfica del árbol obtenido, donde el valor de la variable que aparece en la intersección de cada dos ramas indica el valor de la variable correspondiente a la rama de la izquierda. En los nodos terminales el valor que aparece es el valor que resulta de redondear al valor entero el output medio de los equipos que están incluidos en ese nodo terminal.

De los resultados obtenidos, se puede observar que los valores mas altos de GM corresponden al nodo 2 que está formado por 25 equipos con un valor medio del Output de 2,12, es un grupo homogéneo que no genera ningún nodo hijo. Este grupo está formado por los equipos con mayor índice de diversificación y en él no queda incluido ninguno de los equipos campeones de las ligas analizadas, el 20% de los equipos descendieron de categoría. El segundo nodo en valor del índice GM es el nodo 7 del que forman parte equipos diversificados y en el que se encuentra 2 campeones y solo 1 equipo desciende del total de 27 que conforman el grupo. El grupo del nodo 12 y el del nodo 27 manifiestan valores del índice GM también altos, pero con equipos campeones en el 27 y sin campeones en el 12.

Se observa de estos resultados que no existen patrones claros en la diversificación de la configuración de los equipos que han obtenido los mayores éxitos deportivos. Parece que la máxima diversificación no conduce a los mejores resultados deportivos lo que se podría explicar por la dificultad en generar sinergias de grupo cuando son muchos los miembros del equipo.

Tabla 4

Resumen de resultados de la aplicación de la técnica CART sobre el Modelo 1. Los nodos se numeran de forma que los nodo hijos del nodo n, se identifican como 2n y 2n+1; por ejemplo, los nodos 12 y 13 son hijos del nodo 6

MODELO 1: $OUTPUT = f(GM)$, $loss.matrix = quadratic$

Modelo 1: OUTPUT = f (GM)						Raíz	160	Output Medio	Nodo Terminal
2)						GM>=0.5413133	25	2,1200	*
3)						GM<0.5413133	135		
	6)					GM<0.5399524	108		
		12)				GM>=0.5391956	14	1,9286	*
		13)				GM<0.5391956	94		
			26)			GM<0.538725	82		
				52)		GM>=0.5358893	51	2,3333	*
				53)		GM<0.5358893	31		
					106)	GM<0.5344389	24		
					212)	GM>=0.5290242	8	2,0000	*
					213)	GM<0.5290242	16	2,6250	*
					107)	GM>=0.5344389	7	3,2857	*
			27)			GM>=0.538725	12	3,0000	*
	7)					GM>=0.5399524	27	2,8148	*

Fuente: Elaboración propia.

Tabla 5

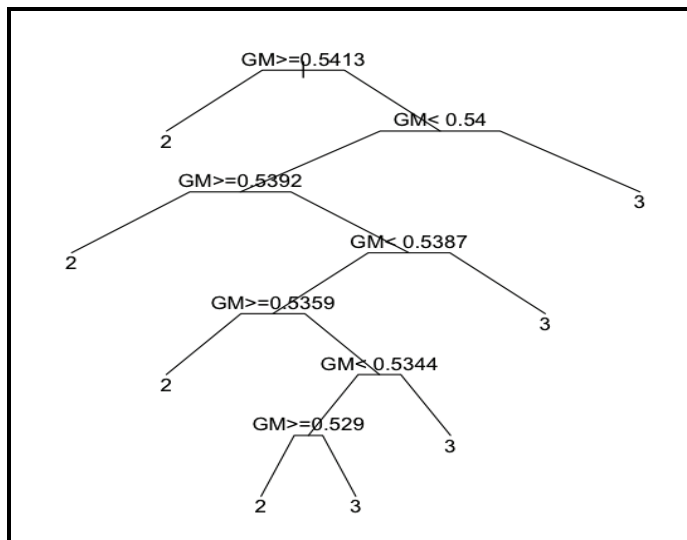
Estadísticos de los contrastes de la estimación CART, para el Modelo 1
(rel. error=1-R²; xerror = Σ error²; xstd= variación de la predicción)

	CP	nsplit	rel error	Xerror	Xstd
1	0.04619565	0	1.0000000	1.0000000	0.10260087
2	0.03260870	2	0.9076087	1.168478	0.09753788
3	0.02173913	5	0.8043478	1.152174	0.09647313
4	0.01000000	7	0.7608696	1.048913	0.09353413

Fuente: Elaboración propia.

Gráfico 1

Representación del resultado de la regresión tree, con la distribución de los equipos y las variables de referencia para el Modelo 1



Fuente: Elaboración propia.

Tabla 6

Distribución de los equipos en los grupos obtenidos de la aplicación de la técnica CART con el Modelo 1

Grupos Modelo 1	Número Equipos	Campeón	2º+3º+4º	5º+6º+7º+8º	9º al 17º	Descienden
2	25		2	4	14	5
12	14		0	2	9	3
52	51	2	5	13	21	10
212	8	1	0	1	3	3
213	16	1	3	3	8	1
107	7	1	3	2	0	1
27	12	1	3	4	4	0
7	27	2	8	3	13	1
	160	8	24	32	72	24

Fuente: Elaboración propia.

Estos resultado pone de manifiesto la necesidad de plantear el modelo sin agregar el valor de sus componentes de forma que se pueda apreciar el efecto se

cada uno de ellos en la generación de los nodos. En la definición del índice los componentes N, S y H permiten identificar los efectos que sobre el índice tienen el número de jugadores, la distribución del tiempo entre ellos y la heterogeneidad de las acciones desarrolladas por los propios jugadores.

La propuesta consiste en considerar los componentes del índice, que tienen sentido económico por sí mismos, como variables en el modelo de clasificación y realizar el análisis a partir de la consideración de estas variables. El modelo en el que se toma como variable dependiente el Output del equipo y como variables explicativas los componentes del índice GM, los factores de número N, de distribución S y de heterogeneidad H, ha dado como resultado el que se presenta en las Tablas 7, 8 y 9, además de la representación gráfica del Gráfico 2.

Los estadísticos de la estimación del modelo se muestran en la Tabla 8. La comparación con la Tabla 5 pone de manifiesto una mejora en los estadísticos del modelo 2 con respecto al modelo 1, lo que es un indicativo de una mejora en la capacidad explicativa de las variables del modelo 2.

Los nodos del árbol en este modelo quedan configurados por el valor de una de las variables determinantes. Se puede observar que la variable H, es la que determina la mayor parte de los nodos generados, por lo tanto es la heterogeneidad en la configuración de los equipos la que determina su pertenencia a uno u otro grupo, el efecto N, que queda determinado por el número de jugadores solo genera la división del nodo 4 en el 8 y el 9, siendo el nodo 8 el que recoge el mayor número de equipos descendidos, el menor output medio y el mayor valor del efecto número de jugadores superando a 28 el número de jugadores de plantilla. Este resultado es coherente con el obtenido en el modelo 1 y pone de manifiesto que un número alto de jugadores con mucha diversidad entre ellos, no genera resultados positivos para el equipo, sin duda por la dificultad en generar sinergias de grupo cuando el grupo es tan numeroso.

El nodo 59 queda configurado por equipos entre los que casi un tercio han quedado entre los 4 primeros y de ellos 3 han sido campeones en las últimas 8 ligas. Se trata de un grupo que queda determinado por equipos con alta diversificación entre ellos y con gran asimetría en la distribución de los tiempos de participación de los jugadores, este grupo identifica a equipos con un grupo de jugadores que participa intensamente en la competición y además con notable heterogeneidad entre los jugadores de la plantilla, lo que podría interpretarse como complementariedad entre los jugadores con más participación y el resto de la plantilla. Además de este grupo, el nodo 15 que queda configurado exclusivamente en base a la variable heterogeneidad, recoge un grupo de equipos donde ninguno baja de categoría y más del 40% están entre los cuatro primeros clasificados. Para este grupo el efecto H toma valores por debajo de la media lo que indica homogeneidad en la plantilla, pero en valores que estarían en el rango medio bajo, podría ser el caso de equipos configurados con jugadores

polivalentes que permiten una adaptación a las circunstancias de la competición sin añadir diversidad a la plantilla.

Tabla 7

Resultados de la aplicación del análisis CART sobre el Modelo 2. Los nodos se numeran de forma que los nodo hijos del nodo n, se identifican como 2n y 2n+1; por ejemplo, los nodos 8 y 9 son hijos del nodo 4

MODELO 2: OUTPUT = f (N, S, H), loss.matrix=quadratic

Modelo 2: OUTPUT= F(N,S,H)					Raíz	160	Output Medio	Nodo Terminal
Nodo								
2)					H<0.0825165	43		
	4)				H>=0.07337072	31		
		8)			N>=0.48245	14	1,4286	*
		9)			N<0.48245	17	2,2941	*
	5)				H<0.07337072	12	2,5833	*
3)					H>=0.0825165	117		
	6)				H>=0.0915095	14	2,3571	*
	7)				H<0.0915095	103		
		14)			H>=0.084116	91		
			28)		S>=-0.0284315	47		
				56)	H>=0.085859	23	2,0435	*
				57)	H<0.085859	24	2,7917	*
			29)		S<-0.0284315	44		
				58)	H<0.0872975	7	2,0000	*
				59)	H>=0.0872975	37	2,7838	*
		15)			H<0.084116	12	3,1667	*

Fuente: Elaboración propia.

Tabla 8

Estadísticos de los contrastes de la estimación CART, para el Modelo 2
(rel. error=1-R²; xerror = $\sum error^2$; xstd= variación de la predicción)

	CP	nsplit	rel error	Xerror	Xstd
1	0.10326087	0	1.0000000	1.0000000	0.10260087
2	0.03396739	1	0.8967391	0.9782609	0.08445799
3	0.01086957	6	0.7228261	0.9565217	0.08320721
4	0.01000000	8	0.7010870	1.0380435	0.08610208

Fuente: Elaboración propia.

Tabla 9

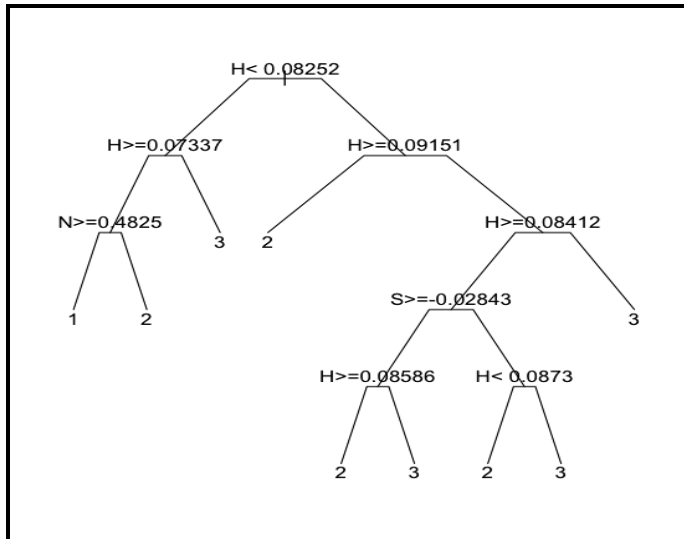
Resultados de la aplicación del análisis CART con identificación de los nodos y de las variables de referencia de cada nodo del Modelo 2

Grupos Modelo 2	Número Equipos	Campeón	2º+3º+4º	5º+6º+7º+8º	9º al 17º	Descienden
8	14		0	1	4	9
9	17	2	2	2	6	5
5	12		2	3	7	0
6	14	1	1	5	3	4
56	23		0	3	18	2
57	24	1	5	7	11	0
58	7		1	1	2	3
59	37	3	9	6	18	1
15	12	1	4	4	3	0
	160	8	24	32	72	24

Fuente: Elaboración propia.

Gráfico 2

Representación del resultado de la regresión tree, con la distribución de los equipos y las variables de referencia para el Modelo 2



Fuente: Elaboración propia.

5. CONCLUSIONES

El análisis realizado en el presente trabajo muestra el potencial de aplicación de las técnicas de análisis complejas en el estudio de los equipos de fútbol profesional. En base a los criterios y políticas de diversificación como herramientas de análisis estratégico se ha planteado el uso del índice de Gollop y Monahan junto con la técnica ‘regression tree’ para comprender y detectar modelos competitivos desarrollados por los equipos que han participado en la Liga española de primera división en las ocho temporadas analizadas. Las variables de referencia son los componentes del índice GM que se obtienen en base al número de jugadores, la distribución del tiempo de participación de los jugadores y las acciones desarrolladas en el terreno.

Se han estimado diversos modelos de los que se deriva que el índice en su conjunto contribuye a explicar los resultados de los equipos, sin embargo, la utilización de los componentes del índice en los modelos permite estudiar con mas detalle las características de los equipos según los resultados alcanzados en la clasificación final.

Los modelos obtenidos muestran que la variable que aparece como la mas determinante de todas en la generación de nodos en los análisis realizados es la variable en la que se mide la heterogeneidad de las acciones del equipo. Es la variable que en todos los modelos genera las particiones que permiten clasificar a los equipos en uno u otro nodo. En segundo lugar aparece la variable que mide la distribución del tiempo de juego entre los miembros de la plantilla.

Parece observarse de forma clara que mucha diversidad en la plantilla puede dar lugar a resultados no deseados, sin duda por la dificultad en generar sinergias de grupo cuando éste es muy numeroso. Cabe destacar que los grupos que generan buenos resultados competitivos son aquellos en los que se observa asimetría en la participación de sus jugadores, con un nivel de diversificación medio alto, y el grupo en el que el nivel de diversificación de los miembros del equipo es media baja.

BIBLIOGRAFIA

- ARCHER, K.J. (2010) “rpartOrdinal: An R Package for Deriving a Classification Tree for Predicting an Ordinal Response”. *Journal of Statistical Software*, 34(7), 1–17.
- BERRY, CH.H. (1971) “Corporate growth and diversification” *Journal of Law and Economics*, 14, 371-383.

- BOON, B. H. & SIERKSMA, G. (2003) "Team formation: Matching Quality Supply and Quality Demand", *European Journal of Operational Research*, 148(2), 277-292.
- BOON, B.H. AND SIERKSMA, G. (2003) "Team formation: Matching quality supply and quality demand", *European Journal of Operational Research*, vol. 148, nº2, pp. 277-292.
- BREIMAN, L., FRIEDMAN, J.H., OLSHEN, R.A. Y STONE, C.J. (1983) *Clasificación and regression trees*. Wadsworth, Belmont, Ca. 1983
- CALIENDO, M. AND RADIC, D (2006) "Ten do it better, do they? An empirical analysis of an Old Football Myth", *IZA Discussion Paper*, nº 2158
- CARMICHAEL, F. AND THOMAS, C. (1995) "Production and efficiency in team sports: an investigation of rugby league football", *Applied Economics*, vol. 27, pp. 859-869.
- CARMICHAEL, F. AND THOMAS, D.A. (2005) "Why did Greece win? An analysis of team performances at Euro 2004", *Management and Management Science Research Institute Working Paper 2006/05*.
- DOBSON, S. AND GODDARD, J. (2003) "Persistence in sequences of football match results: A Monte Carlo analysis", *European Journal of Operational Research*, vol. 148, nº2, pp. 247-256.
- ESPITIA-ESCUER, M. AND GARCÍA-CEBRIÁN, L.I. (2004) "Measuring the Efficiency of Spanish First-Division Soccer Teams", *Journal of Sports Economics*, vol. 5, pp. 329-346.
- ESPITIA-ESCUER, M. AND GARCÍA-CEBRIÁN, L.I. (2006) "Performance in Sports teams: results and potential in the professional Soccer league in Spain", *Management Decision*, vol. 44, pp. 1020-1030.
- ESTEBAN GARCÍA J. Y COLL SERRANO V. (2003) "Competitividad y Eficiencia" *Estudios de Economía Aplicada*, vol 21-3, pp. 423-450.
- GARCÍA J., AND RODRÍGUEZ P. (2002) "The Determinants of Football Match Attendance Revisited", *Journal of Sports Economics*, vol. 3, nº 1.
- GOLLOP, F. M. AND MONAHAN, J.L. (1991): "A generalized index of diversification: Trends in U.S. Manufacturing", *The Review of Economics and Statistics*, vol. 73, nº 2, pp. 318-330.
- GORT M. (1962) *Diversification and Integration in American Industry*. (Princeton; Princeton Univ. Press.
- GRAHAM, J. R., LEMMON M., AND WOLF J., (2002) "Does Corporate Diversification Destroy Value?", *Journal of Finance* 57.
- HOSKISSON R., HITT M. A., JOHNSON R. A., MOESEL DOUGLAS D. (1993) "Construct validity of an objective (entropy) categorical measure of diversification strategy", *Strategic Management Journal*, vol. 4, nº 3.
- JACQUEMINE A. & BERRY Ch. (1979), "Entropy measurement of Diversification and Corporate Growth." *The Journal of Industrial Economics* 27, pg. 359-369.

- KLAASSEN, F. J. G. M & MAGNUS, J. R. (2003) "Forecasting the Winner of a Tennis Match". *European Journal of Operational Research*, 148(2), 257-267.
- KONING, R.H., KOOLHAAS, M., RENES, G. AND RIDDER, G. (2003) "A simulation model for football championship", *European Journal of Operational Research*, vol. 148, nº2, pp. 268-276.
- LEWIS, A.J. (2005) "Towards fairer measures of player performance in one-day cricket" *Journal of the Operational Research Society*, vol. 56, nº 7, pp. 804-815.
- LEWIS, R. J. (2000) "An Introduction to Classification and Regression Tree (CART) Analysis". Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California.
- MARKIDES C. C. AND WILLIAMSON P. J., (1994) "Related diversification, core competences and corporate performance", *Strategic Management Journal*, vol. 15.
- MOSCHINI, G.C. (2004) "Nash equilibrium in strictly competitive games: live play in soccer", *Economics Letters*, vol. 85, nº 3, pp. 365-371.
- PEHRSSON, A. (2006) "Business relatedness measurements: State-of-the-art and a proposal". *European Business Review*. Vol. 18, n. 5, pg. 350 – 363
- RAJAN R., SERVAES H., ZINGALES L., (2000) "The Cost of Diversity: The Diversification Discount and Inefficient Investment", *Journal of Finance* 55.
- TROUTT M. D., ACAR M. (2005) "A Lorenz–Pareto measure of pure diversification", *European Journal of Operational Research*, vol 167, nº 2.