



Actividad Colaborativa Módulo 4

Pablo Olmos

ÍNDICE

- 1.-Definición del Objetivo, en este caso la métrica.
- 2.-Fuente de Datos con la que vamos a trabajar
- 3.-Desarrollo de la Métrica
- 4.-Contraste y Validez
- 5.-Comentarios personales.

1.-Definición de la métrica.

Se nos plantea idear y llevar a cabo una métrica que sea aplicable a mundo del deporte. Hemos pensado desarrollar, dentro del fútbol, una métrica que nos permita predecir cuál será la clasificación de un equipo dentro de la Liga española.

Nuestra idea es desarrollar de forma evolutiva esta idea y que nos permita implementar mejoras en nuestra estrategia deportiva a lo largo de la temporada, por lo que puede ser una herramienta importante para la detección de carencias y desarrollo de virtudes dentro de nuestro equipo.

2.-Fuente de Datos.

Vamos a trabajar con el Fichero de Opta laliga201920.csv en el cual aparecen los datos de los jugadores a final de temporada. La idea original es implementar este desarrollo de forma dinámica y semanal en función de la recogida de información.

3.-Desarrollo de la Métrica

Las variables de Opta las vamos a agrupar y suma a nivel Equipo. Así obtenemos una visión global de todos los integrantes de la plantilla.

El nombre de la métrica será Score_Clasificacion y dependerá de 5 grandes factores que hemos considerado determinantes en el juego, estos son:

Factor1. Contundencia

El objetivo es medir la capacidad de salir victorioso en la disputa del balón, en todas sus modalidades. Para ello necesitamos las siguientes variables y fórmulas.

Para ello calculamos 2 variables auxiliares

$$['duelos_ganados'] = ['duels_ground_won'] + ['duels_aerial_won']$$
$$['duelos_total'] = ['duels_ground_total'] + ['duels_aerial_total']$$

Que nos ayudaran a construir el factor.

$$['contundencia'] = ['duelos_ganados'] / ['duelos_total']$$

Factor2. Asociación

El objetivo es medir la capacidad de asociación de nuestro equipo mediante el pase exitoso en todas sus modalidades.

Para ello calculamos 2 variables auxiliares

$$['paseok'] = ['passes_short_successful'] + ['passes_long_successful']$$
$$['pasetotal'] = ['passes_short_successful'] + ['passes_short_unsuccessful'] +$$
$$['passes_long_successful'] + ['passes_long_unsuccessful']$$

Que nos ayudaran a construir el factor

$$['asociacion'] = ['paseok'] / ['pasetotal']$$

Factor3. Acierto Cara Puerta

Muy sencillo simplemente a partir de dos variables medimos el rendimiento goleador de nuestro equipo en función del total de disparos.

$$['Acierto_Cara_Puerta'] = ['goals'] / ['shots_total']$$

Factor4. Capacidad de Dribbling y Asistencia

Con este factor pretendemos medir la capacidad de dribbling de nuestros jugadores asociada a la capacidad de brindar asistencias. Partimos de la idea de que existe una relación muy importante entre ambas variables.

Para ello calculamos 2 variables auxiliares

$['dribSuccess'] = ['dribbles_successful'] / (['dribbles_unsuccessful'] + ['dribbles_successful'])$

$['Asistencia'] = (['keypasses'] + ['assists'] + ['assists_intentional'] + ['assists_second_goal']) / ['passes']$

Y a partir de estas 2 variables, una el éxito el dribbling y otra el nivel de asistencias sobre el total de pases, generamos nuestro Factor

$['dribIntention'] = ['dribSuccess'] + (['Asistencia'])$

Factor5. Great GoalKeeper

Generamos una factor combinado que nos mida la capacidad de nuestro portero en función a 2 ejes, uno la capacidad para parar los disparos a puerta y otro el juego con los pies.

Para ello calculamos 2 variables previas

$['intervenciones'] = ['clearances'] + ['interceptions'] + ['blocks']$

$['indice_paradas'] = 1 - (['goals_conceded'] / ['intervenciones'])$

El índice de paradas nos habla de los goles recibidos en función de las intervenciones, esto lo realizamos para ponderar al típico portero que disparan 30 veces en un partido contra otro al que disparan 5.

Y la capacidad de jugar con los pies de nuestro portero de forma exitosa

$['pasesokgk'] = ['passes_open_play_successful'] / ['passes_open_play']$

A partir de estas 2 vamos a crear el Factor Great Goalkeeper dándole diferentes pesos a cada una de ellas. Consideramos que el índice de paradas pesará un 80% sobre la nota y el juego con los pies un 20%. De lo que tenemos:

$['Geat_GK'] = 0.8 * ['indice_paradas'] + 0.2 * ['pasesokgk']$

4.-Contraste y Validez

A partir de estos 5 factores y dándoles la misma importancia en el juego de nuestro equipo crearemos el factor final:

$$\text{Score_Clasificacion} = 1/5 * [\text{'contundencia'}] + 1/5 * [\text{'asociacion'}] + 1/5 * [\text{'Acierto_Cara_Puerta'}] + 1/5 * [\text{'driblention'}] + 1/5 * [\text{'Geat_GK'}].$$

Una vez obtenida esta nota, rankeamos de 1 a 20 nuestro Score y contrastaremos con la clasificación final de la liga.

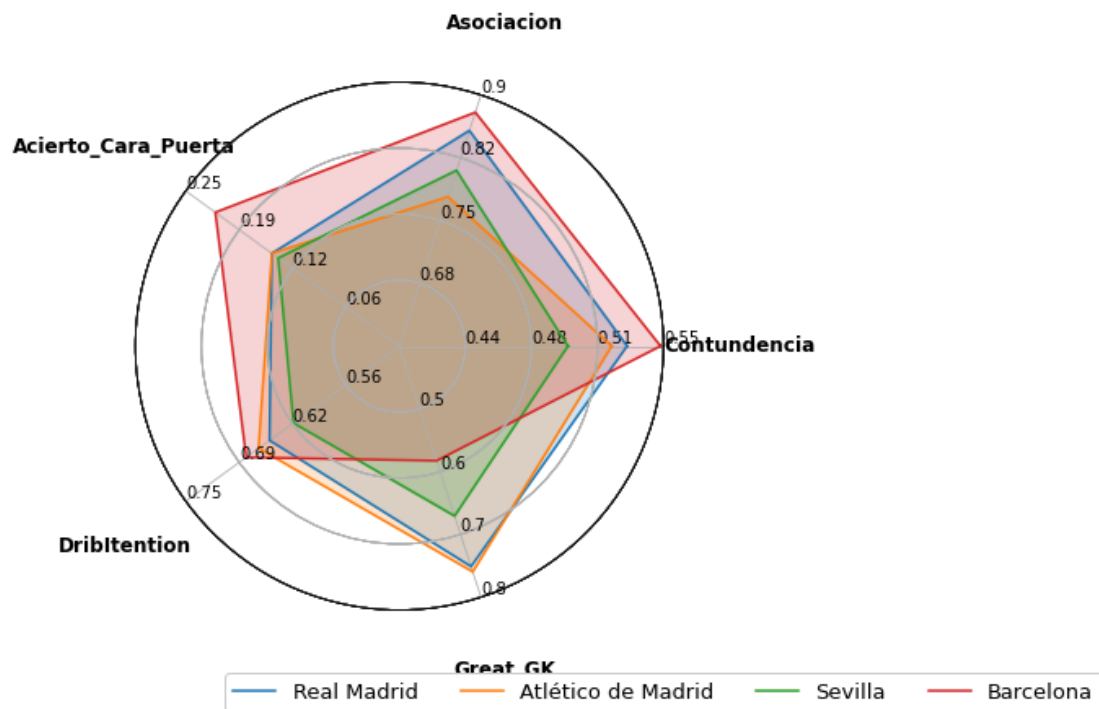
| | Clasificación_Real | Score_Clasificacion |
|--------------------|--------------------|---------------------|
| Real Madrid | 1 | 1 |
| Barcelona | 2 | 2 |
| Atlético de Madrid | 3 | 3 |
| Sevilla | 4 | 4 |
| Villarreal | 5 | 5 |
| Real Sociedad | 6 | 7 |
| Granada CF | 7 | 13 |
| Getafe | 8 | 15 |
| Valencia CF | 9 | 6 |
| Osasuna | 10 | 14 |
| Athletic Club | 11 | 9 |
| Levante | 12 | 11 |
| Real Valladolid | 13 | 12 |
| Eibar | 14 | 19 |
| Real Betis | 15 | 10 |
| Alavés | 16 | 20 |
| Celta de Vigo | 17 | 8 |
| Leganés | 18 | 16 |
| Mallorca | 19 | 18 |
| Espanyol | 20 | 17 |

Parece que en los 5 primeros hemos acertado plenamente.

Nos llama mucho la atención lo bien que hemos considerado al Celta de Vigo en contraste con su clasificación liguera, también muy curioso el contraste para el Getafe, el Score lo considera como mucho peor equipo de lo que la realidad demuestra.

5.-Comentarios personales.

Visualización Radar Chart del Top 4 de la Liga

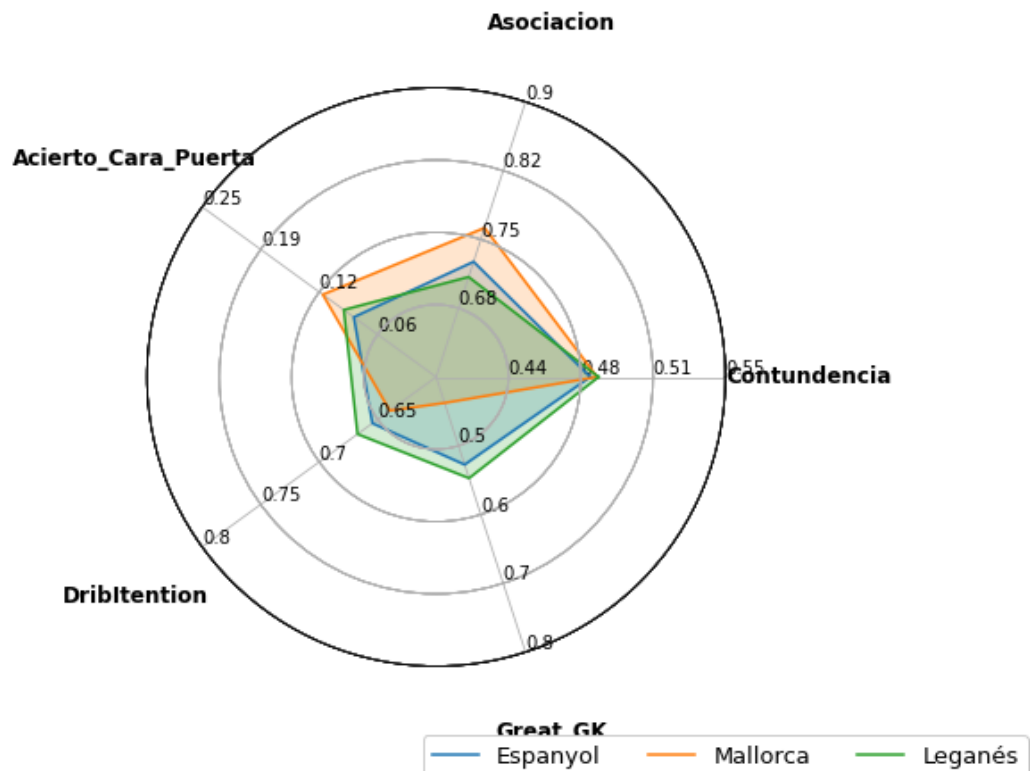


Parece que según nuestro Score, el Barcelona es el equipo que mejor contrala todas las facetas del juego incluso en la Contundencia en los balones disputados por encima del Atlético de Madrid pero le ha penalizado mucho su falta de solidez defensiva tal y como se refleja en el gráfico.

El Real Madrid parece sin duda el equipo más equilibrado, destaca también en el gráfico la escasa capacidad de asociación en el juego del Atlético de Madrid.

Gran estadística defensiva del Real Madrid, la del Atlético de Madrid no llama tanto la atención ya que intuitivamente es un valor esperado.

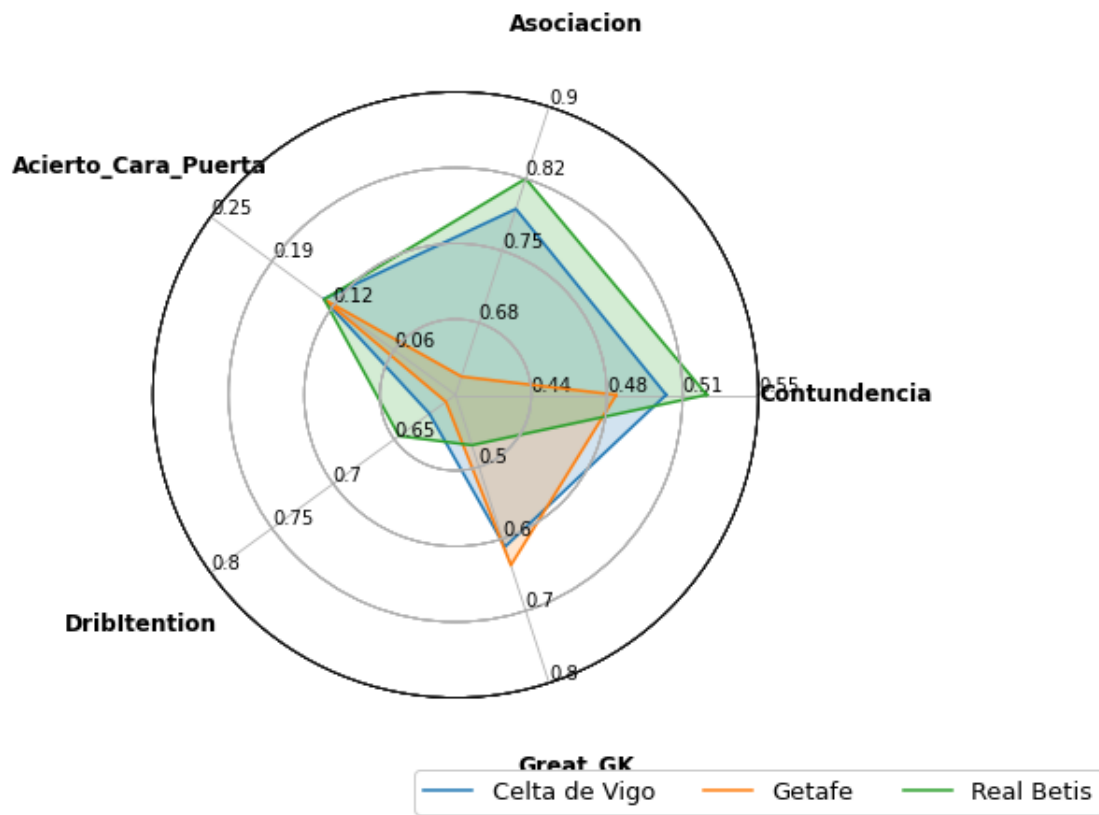
Visualización Radar Chart del Top 3 por la cola de la Liga.



Muy Poco destacables las estadísticas de los equipos descendidos como era de esperar, quizá sorprende el Mallorca ya que propone un fútbol de asociación y tiene cierta capacidad goleadora. Pero defensivamente ha sido muy penalizado.

De estos equipos llama la atención la falta de contundencia en el juego que a priori parece la virtud de los equipos de la parte baja de la tabla.

Y por último vamos a representar a los equipos que nos han sorprendido la diferencia entre la clasificación estimada por el Score y la clasificación real. Son el Getafe, el Celta de Vigo y el Real Betis.



Llama mucho la atención tanto en el Celta de Vigo como en el Real Betis que ambos equipos tienen una gran capacidad de asociación en el juego pero al mismo tiempo no se materializa en driblings y asistencias que son claves para el Gol. Quizá sean equipos de toque pero sin profundidad.

En la otra punta nos encontramos con el Getafe, que propone un juego directo y con una muy fuerte solidez defensiva.

Por último como a desarrollar en el futuro y viendo la importancia de las métricas defensivas proponemos desarrollar más esta parte y/o ponderar la que tenemos de forma diferente ya que parece que es fundamental en el desarrollo de la clasificación ligera.