

Big Data y el Análisis de Conglomerados: Descubriendo los Diferentes Perfiles de los Mediocentros – Informe

Para la realización del estudio se plantea lo siguiente. El centro del campo es un lugar donde se abarcan casi todos los parámetros que medimos de los jugadores donde muchos de ellos no tienen un perfil específico como podría ser por ejemplo el de un mediocentro defensivo. Es por ello por lo que en este estudio se intenta identificar y diferenciar a los diferentes mediocentros puro según sus datos.

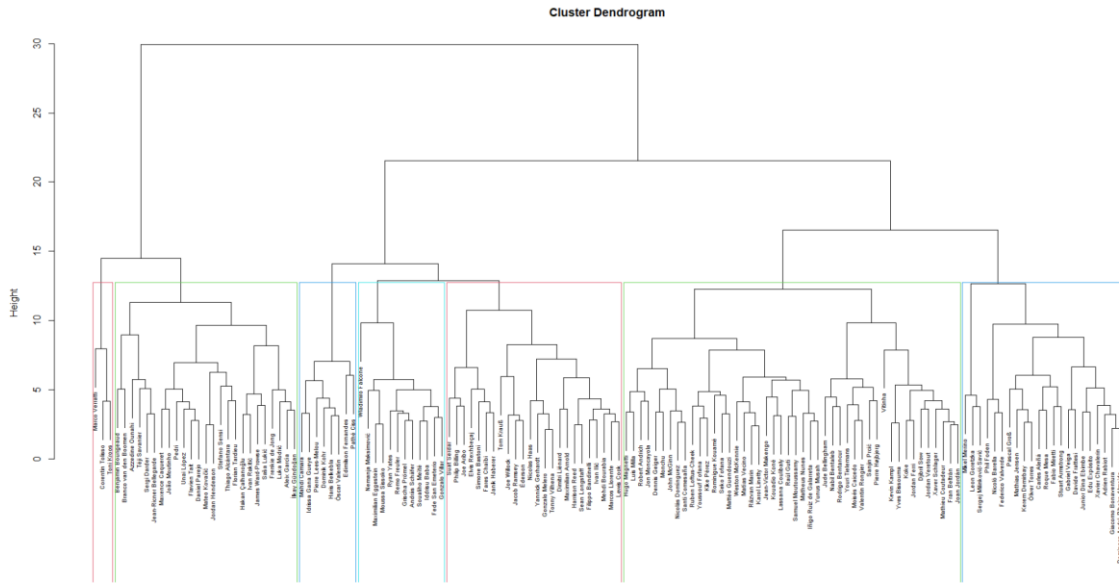
Para ello se ha obtenido diferente información de la web de datos abiertos FBREF para los jugadores que sean mediocentros (información de Transfermarkt) que hayan jugado un mínimo del 40% de minutos esta temporada. Se ha dividido todo en tres grandes bloques para poder crear una valoración a partir de un algoritmo para una mejor visualización de los datos. Las variables recogidas son las siguientes.

- Defensivo: Intercepciones, entradas, veces regateado, bloqueos, despejes, porcentaje y cantidad de duelos aéreos ganados.
- Creación de juego: Pases cortos y medios (realizados y porcentaje de acierto), pases largos realizados, pases progresivos, toques, recepciones progresivas.
- Ofensivo: Regates realizados y porcentaje de acierto, asistencias esperadas, pases clave, pases en el último tercio, tiros a puerta, goles, goles esperados, acciones de tiro creadas.

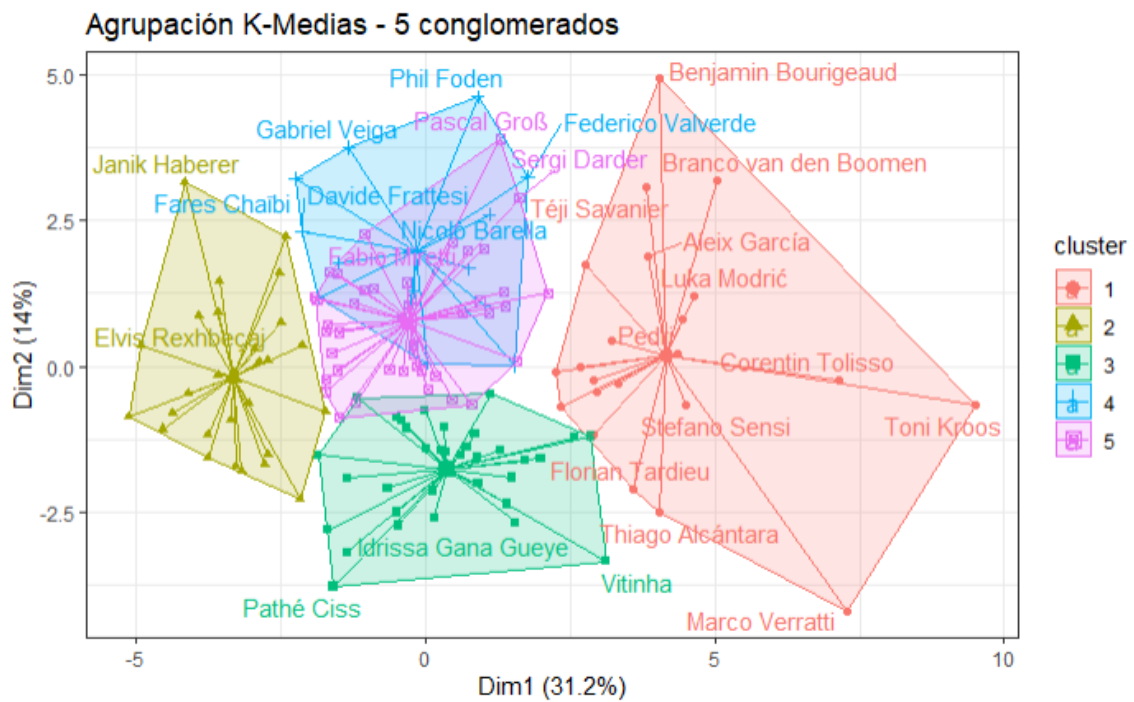
Todos los datos han sido normalizados a 90 minutos.

Una vez tenemos toda la información recogida y normalizada para poder utilizarla se procede a el análisis por conglomerados. Previo a la aplicación de este se debe escoger primero el número de clusters (grupos) que se realizarán. Usando la librería NbClust encontramos como la mayor cantidad de veces se recomienda usar 5 clusters.





Visto que la mejor separación es en 5 grupos procedemos a realizar el análisis por conglomerados. A continuación, encontramos los jugadores graficados en las primeras dos componentes del proceso del ACP agrupados según cluster.



Ahora que ya conocemos la cantidad de perfiles vamos a ver las cualidades de cada uno de ellos. Para conocer de forma más rápida y visual como se dividen los clusters enseñaremos el Score promedio de cada grupo en los tres aspectos observados.

Cluster	Attack Score	Creation Score	Defensive Score	Promedio Scores
1	0,6709	0,6508	0,3562	0,5593
2	0,2779	0,2205	0,4933	0,3305
3	0,3905	0,4894	0,5224	0,4674
4	0,6457	0,3548	0,5401	0,5135
5	0,4981	0,3727	0,3659	0,4123

Podemos apreciar como hay unos perfiles muy marcados. Por ejemplo, el segundo cluster sería el de un perfil más defensivo el cual tal vez podría reconvertirse a un mediocentro defensivo. Otros como el primero podrían ser más típicos de un mediocentro ofensivo o de banda donde aporta de forma importante tanto en el ataque como en la creación de juego dejando un poco atrás la faceta defensiva. Además, vemos el promedio de los tres scores, siendo el primer y el cuarto grupo los que abarcan más variables, los que se le podría llamar más “todo campistas”.

Ahora, una vez visto esto, acudimos a Transfermarkt para ver algunos de los mediocentros de la liga española con mayor valor de mercado para ver en que perfil encajarían y que valores obtienen en los diferentes scores respecto a los de su grupo.

Player	Grupo	Attack Score	Diferencia Attack Score	Creation Score	Diferencia Creation Score	Defensive Score	Diferencia Defensive Score
Marcos Llorente	5	0.49	-0.01	0.33	-0.04	0.52	0.15
Frenkie de Jong	1	0.75	0.08	0.69	0.04	0.5	0.15
Pedri	1	0.75	0.08	0.66	0.01	0.16	-0.2
Aleix García	1	0.87	0.2	0.61	-0.04	0.48	0.13
Toni Kroos	1	0.82	0.15	1	0.35	0.25	-0.11
Federico Valverde	4	0.85	0.2	0.51	0.16	0.34	-0.2
Mikel Merino	4	0.65	0.01	0.39	0.04	1	0.46
Yunus Musah	5	0.35	-0.15	0.3	-0.08	0.46	0.09

Vemos como salvo algún outlier como el rendimiento defensivo de Mikel por ejemplo, los valores son bastante cercanos a los que se le adjudicaron en su grupo por lo que el objetivo de crear grupos parece haberse logrado.

En resumen, el análisis de datos y el análisis por conglomerados son herramientas valiosas para entender mejor el mundo del fútbol y tomar decisiones informadas. Este estudio es solo un ejemplo de cómo el Big Data está ayudando a mejorar nuestra comprensión del juego y a identificar patrones y tendencias en el rendimiento de los jugadores.