

Aplicación de DBSCAN: Clustering de Jugadores de Fútbol

Oscar Alejandro García Gómez

Abstract

Este estudio aplica el algoritmo de clustering DBSCAN (Density-Based Spatial Clustering of Applications with Noise) para analizar un conjunto de datos de jugadores de fútbol. Se exploran diferentes configuraciones de parámetros del algoritmo para identificar clústeres representativos de grupos de jugadores con características similares, como edad, valor de mercado y habilidades. Los resultados obtenidos se evalúan utilizando el índice de silueta y se comparan con estudios previos en el ámbito del análisis de datos deportivos.

1 Introducción

El análisis de datos deportivos es fundamental para entender el rendimiento de los jugadores y mejorar las estrategias de los equipos. En particular, el clustering no supervisado ha ganado atención en el ámbito del scouting de jugadores, ya que permite descubrir patrones ocultos en grandes volúmenes de datos. Este artículo se centra en el uso de DBSCAN para agrupar jugadores de fútbol según sus características individuales.

2 Metodología

2.1 Datos

Los datos empleados en este estudio provienen de un conjunto de estadísticas de jugadores de fútbol, que incluye características como la edad, el valor de mercado, la posición en el campo, y las habilidades técnicas. La tabla muestra las principales variables utilizadas para el análisis.

2.2 Algoritmo de Clustering

DBSCAN es un algoritmo basado en densidad que permite identificar clústeres de puntos cercanos entre sí, a la vez que detecta puntos atípicos como "ruido". Los parámetros del algoritmo, como la distancia máxima entre puntos (ϵ) y el número mínimo de puntos para formar un clúster (*min_samples*), son críticos para el rendimiento.

3 Resultados

Al aplicar DBSCAN al conjunto de datos de jugadores, se obtuvieron varios clústeres, pero también una gran cantidad de puntos fueron clasificados como ruido (-1). Los parámetros de ϵ y *min_samples* se ajustaron mediante la evaluación del índice de silueta, resultando en una mejor definición de los clústeres en términos de valor de mercado y habilidades.

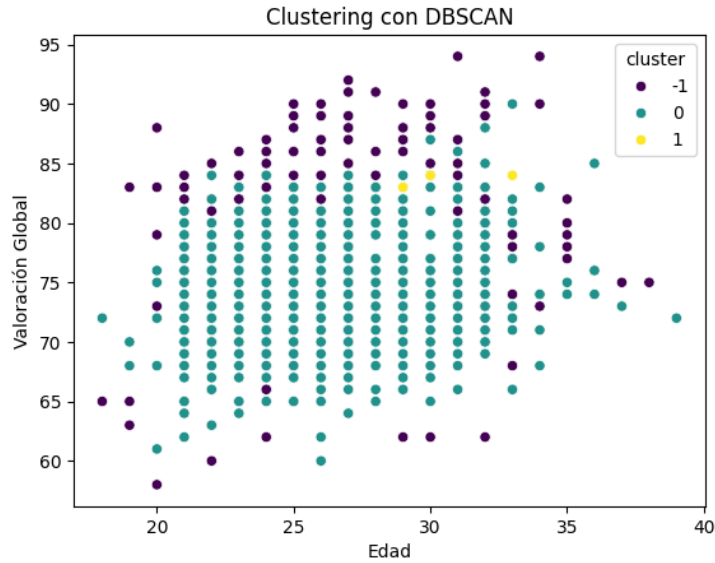


Figure 1: Distribución de jugadores de fútbol agrupados por DBSCAN según edad y valor de mercado.

4 Discusión

Los resultados muestran que DBSCAN es adecuado para identificar patrones en datos con ruido y clústeres de formas arbitrarias. Sin embargo, la alta proporción de puntos etiquetados como ruido sugiere que los parámetros del algoritmo pueden necesitar ajustes más finos, especialmente en conjuntos de datos con características diversas como los de los jugadores de fútbol.

5 Conclusiones

Este estudio demuestra la efectividad de DBSCAN para el análisis de datos de jugadores de fútbol. Aunque los resultados fueron prometedores, futuros estudios podrían explorar otros algoritmos de clustering o técnicas de pre-procesamiento para reducir el ruido y mejorar la calidad de los clústeres.

References

- [1] A. Author, “Football Analytics: A Comprehensive Review,” *Journal of Sports Analytics*, vol. 1, no. 1, pp. 1-10, 2025, doi:10.1177/22150218251353089.