

Clasificación de Jugadores de FC25 mediante K-means y Análisis de Componentes Principales

Samuel Efrén Castro Arias. Gerónimo Martínez Ávila, Dixon Merchan Gutierrez

October 7, 2024

1 Introducción

Este trabajo presenta un análisis de agrupamiento no supervisado de jugadores del popular videojuego FC25 (FIFA) utilizando el algoritmo K-means. El objetivo es identificar cuatro clusters óptimos de jugadores basados en sus características, empleando distintas métricas de distancia y visualizando los resultados mediante Análisis de Componentes Principales (PCA).

2 Descripción del Conjunto de Datos

Se realizó un análisis exploratorio de datos (EDA) sobre un conjunto de datos de FC25 previamente organizado. Las etapas del EDA incluyeron:

- * *Limpieza:* Eliminación de variables no numéricas y datos faltantes.
- * *Codificación:* Conversión de la variable "pie preferido" a una representación binaria.
- * *Imputación:* Rellenado de los valores faltantes utilizando técnicas adecuadas.

3 Metodología

3.1 Algoritmo K-means

El algoritmo K-means es una técnica de agrupamiento que busca particionar un conjunto de datos en k grupos (clusters) de modo que los elementos dentro de cada grupo sean lo más similares posible, mientras que los elementos de diferentes grupos sean lo más distintos posible.

3.1.1 Selección del número de clusters

Se empleó el método del codo para determinar el valor óptimo de k , es decir, el número de clusters que mejor representa la estructura subyacente de los datos.

3.1.2 Métricas de distancia

Se utilizaron tres métricas de distancia para evaluar la similitud entre los puntos de datos:

- * *Distancia de Mahalanobis:* Considera la correlación entre las variables.
- * *Distancia Euclidiana:* Mide la distancia "en línea recta" entre dos puntos en un espacio euclidiano.
- * *Distancia L1 (Manhattan):* Suma las magnitudes de las diferencias por coordenadas.

3.1.3 Implementación

Se implementó el algoritmo K-means utilizando librerías especializadas, así como una implementación manual para explorar las distintas métricas de distancia.

3.2 Análisis de Componentes Principales (PCA)

Para visualizar los resultados de la agrupación en un espacio de menor dimensión, se aplicó PCA. Se seleccionaron cuatro componentes principales que explicaban el 80

4 Resultados

Los resultados del análisis mostraron que el algoritmo K-means, combinado con las diferentes métricas de distancia y la visualización mediante PCA, permitió identificar cuatro clusters de jugadores con características distintivas.

5 Conclusiones

Este estudio demuestra la utilidad del algoritmo K-means para clasificar jugadores de FC25 en grupos homogéneos. La elección de la métrica de distancia adecuada y la aplicación de PCA resultaron fundamentales para obtener una visualización clara y comprensible de los resultados.