

Minería de Datos

Lab. #7

Ivan Saavedra, Ph.D.

saavedrai@uninorte.edu.co

Universidad del Norte
División de Ingenierías
Dpto. Ingeniería de Sistemas



202110

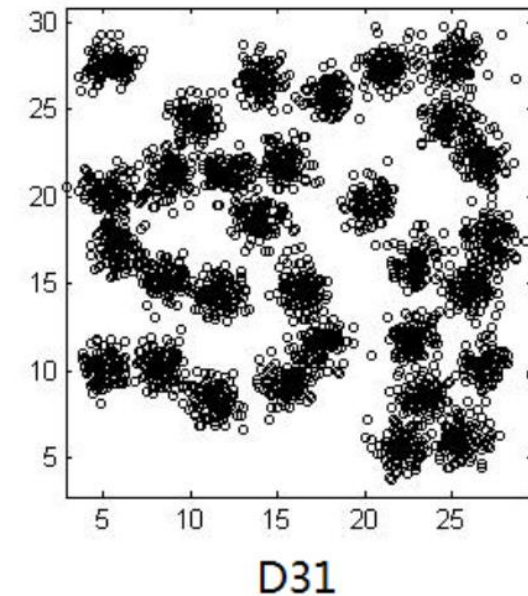
1. Clustering D31 Dataset

Aplice las técnicas de clustering K-means y DBSCAN para encontrar grupos de datos en el dataset suministrado (D31_data.xlsx fuente: <http://cs.joensuu.fi/sipu/datasets/>).

C.J. Veenman, M.J.T. Reinders, and E. Backer, A maximum variance cluster algorithm. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2002. 24(9): p. 1273-1280.

Metodología

- Definir parámetros de los modelos
- Ajustar el modelo
- Compare los resultados de las dos técnicas y mencione cual seria la mejor para este caso



2. Clustering Fifa Players

Aplice K-means clustering en el dataset suministrado de jugadores Fifa con el objetivo de encontrar grupos de jugadores que presenten similitudes y así proveer múltiples opciones a la hora de obtener nuevos jugadores en un equipo de futbol.

Metodología

- Preparar el dataset
- Definir parámetros del modelo
- Ajustar el modelo
- Evaluar el modelo
- Realizar este ejercicio
 1. Utilizando los siguientes features (Age, Potential)
 2. Utilizando los siguientes features (Age, Overall, Potential, Value, Skill Moves, Height, Weak Foot)



<https://www.transfermarkt.com/>

Clustering

Entregable:

- Un archivo de Jupyter Notebook con el desarrollo del análisis
- Las conclusiones y respuestas al objetivo del análisis deben ser contestadas en el mismo notebook.
- Se sugiere que comente las secciones de manera adecuada para una mejor interpretación de su análisis.
- La fecha de entrega es el día 17 de Mayo del 2021 vía catalogo web enlace de laboratorios.