

Actividad 4 – Aprendizaje No Supervisado

Brayan Steven Bonilla Castellanos
Juan Carlos Monsalve Gómez

Universidad Ibero.
Ingeniería de Software
Inteligencia artificial

Enlace repositorio GitHub

https://github.com/Brayan-Bonilla1224/IA_Unidad3_act4

Enlace video-tutorial

<https://youtu.be/WFJOB2Kd224>

Descripción del ejercicio

Escribir en Python un ejercicio que cumpla con el modelo de aprendizaje no supervisado.

Ejecución del ejercicio

Importamos las librerías necesarias para la ejecución del programa

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

Establecemos una semilla de reproducibilidad con un valor de 42, luego generamos un arreglo comprendido en el rango de 10 a 100 para un valor de 2 cifras, con un tamaño de 30 posiciones para los valores de X y Y

```
np.random.seed(42)

X = np.random.randint(10, 100, size=30)
Y = np.random.randint(10, 100, size=30)
```

Creamos un DataFrame a partir de los valores obtenidos anteriormente, validamos la creación imprimiendo por consola, luego procedemos a pintar la gráfica de dispersión

```
data = pd.DataFrame({'X': X, 'Y': Y})
print(data)

plt.scatter(data['X'], data['Y'])
plt.show()
```

Utilizamos el algoritmo de KMeans el cual nos permite definir un numero de clusters centrales y entrenar con la data obtenida del DataFrame, luego imprimimos los resultados obtenidos (Ubicaciones en el grafo)

```
kmeans = KMeans(n_clusters=5).fit(data)
centroids = kmeans.cluster_centers_
print(centroids)
```

Por último, creamos el grafo de dispersión el cual nos va a mostrar la información obtenida anteriormente, añadiendo la ubicación de los centroides de color rojo, identificando claramente la agrupación que identifico el sistema basado en el algoritmo KMeans

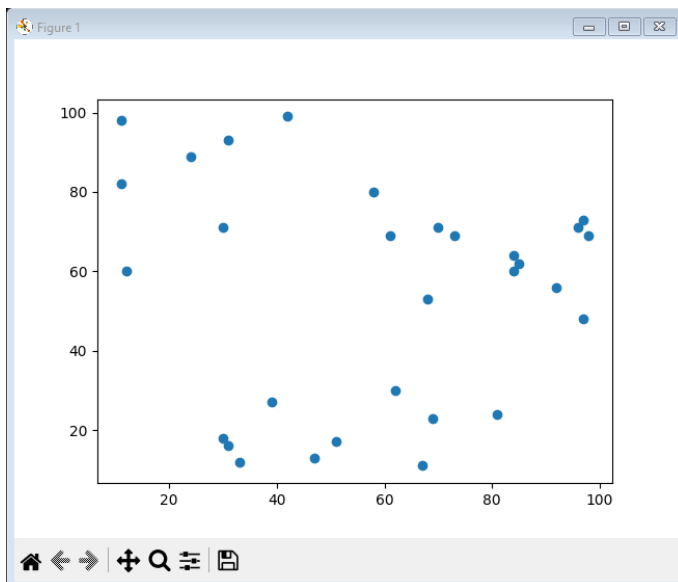
```
plt.scatter(data['X'], data['Y'], c=kmeans.labels_.astype(float), s=50, alpha=0.5)
plt.scatter(centroids[:,0], centroids[:,1], c='red')
plt.show()
```

Ejecución del ejercicio

Obtenemos los valores del DataFrame

	X	Y
0	61	69
1	24	89
2	81	24
3	70	71
4	30	71
5	92	56

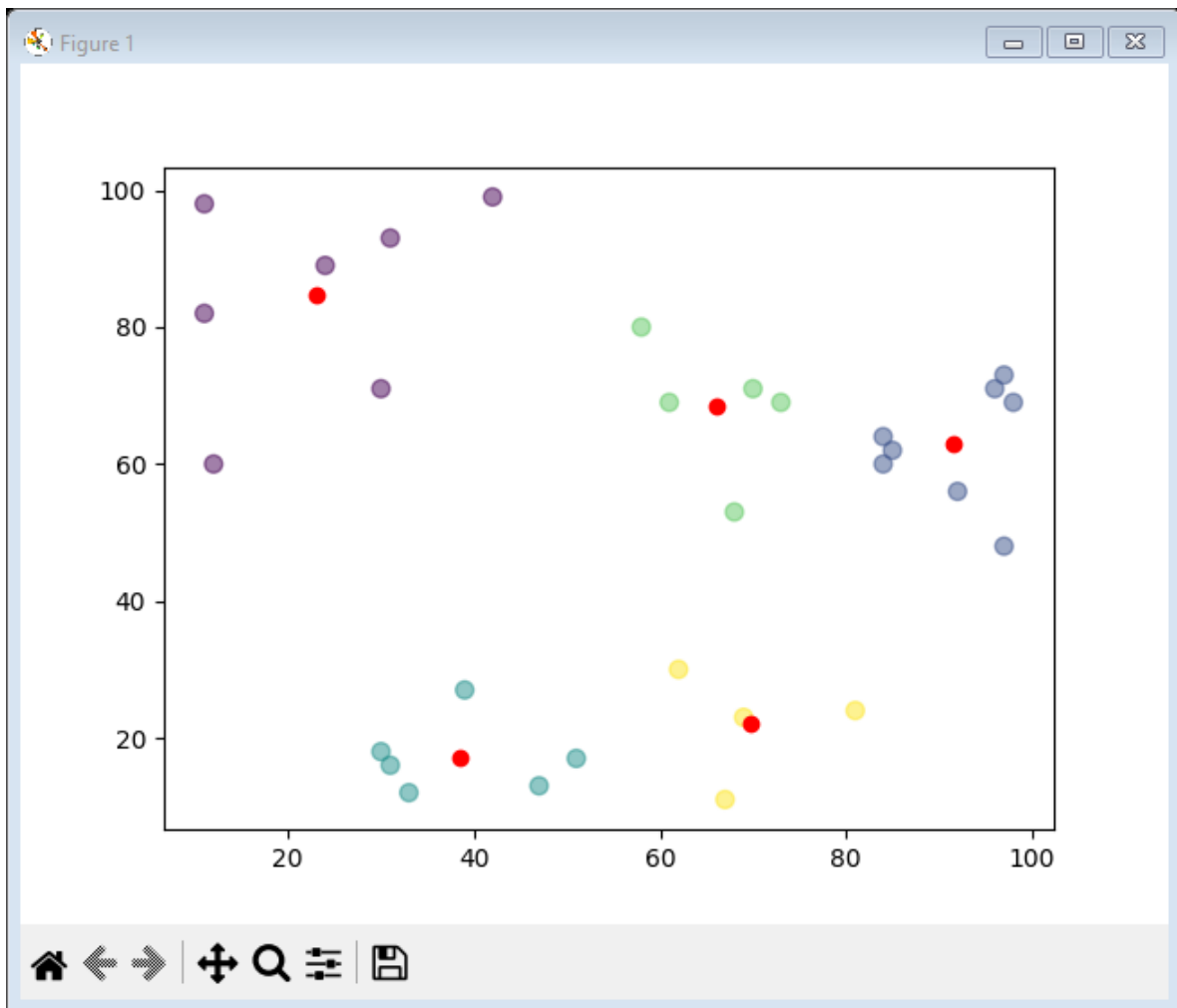
Dibujamos la gráfica de dispersión obtenida con los valores del DataFrame



Visualizamos la ubicación de los centroides, en donde denota la agrupación que identifico el sistema, basado en el algoritmo KMeans

```
[[23.      84.57142857]  
 [91.625   62.875    ]  
 [38.5     17.16666667]  
 [66.      68.4      ]  
 [69.75    22.       ]]
```

Por último, evidenciamos la agrupación realizada por el método de aprendizaje no supervisado, en donde los centroides identificados de color rojo agrupan sus datos con un color específico, de tal forma nos permite realizar análisis y toma de decisiones



Conclusiones

- La técnica de aprendizaje no supervisado es muy útil, por ejemplo, para encontrar y explorar patrones y estructuras de datos inmersas en un conjunto de datos sin etiquetar.
- El uso de las técnicas de aprendizaje no supervisado va mucho más allá de una simple estadística, llegando a campos tan diversos como el análisis de segmentos de clientes en grupos similares según comportamientos de los mismos, el análisis de anomalías en sistemas de Ciberseguridad, redes eléctricas, etc.
- Todo análisis de datos parte de unos objetivos, siendo entonces lo más importante el diseño de la investigación o la estrategia de análisis, no la herramienta, es fundamental tener en cuenta que la herramienta es solo un medio.

Bibliografía

¿Qué es el aprendizaje no supervisado? / IBM. (s. f.). <https://www.ibm.com/es-es/topics/unsupervised-learning>