



Actividad 4 - Métodos de aprendizaje no supervisado

Jessica Carolina Osorio Hernández (ID 100112053 / josori10@ibero.edu.co)

Duván Arley Hernández Niño (ID 100110002 / dherna84@ibero.edu.co)

Corporación Universitaria Iberoamericana
Ingeniería de Software
Ingeniero: Jorge Castañeda

Sasaima Cundinamarca
Abril 2024

Documento con la descripción de los datos:

- **Estacion_Inicio:** La estación donde comienza el viaje. Es un número entero aleatorio entre 1 y 7.
- **Estacion_Fin:** La estación donde termina el viaje. Es un número entero aleatorio entre Estacion_Inicio+1 y 8.
- **Hora_Inicio:** La hora del día cuando comienza el viaje. Comienza a las 8:00 AM y se incrementa en 15 minutos para cada viaje.
- **Hora_Fin:** La hora del día cuando termina el viaje. Se calcula sumando la Duración a la Hora_Inicio.
- **Duracion:** La duración del viaje en minutos. Se calcula como 10 minutos por cada estación que el tren debe pasar.
- **Capacidad_Tren:** La capacidad del tren. Es un número entero aleatorio entre 80 y 120. - Pasajeros: El número de pasajeros en el tren. Es un número entero aleatorio entre 0 y Capacidad_Tren.
- **Cluster:** El grupo asignado al viaje por el algoritmo de agrupamiento K-means.

Explicación del aprendizaje no supervisado:

El aprendizaje no supervisado es un tipo de aprendizaje automático que se utiliza para extraer patrones de un conjunto de datos sin tener etiquetas de salida o una retroalimentación específica. En lugar de tener un objetivo claro, el aprendizaje no supervisado se utiliza para encontrar estructuras ocultas y relaciones en los datos.

En este caso, estamos utilizando el algoritmo de agrupamiento K-means, que es una técnica de aprendizaje no supervisado. Este algoritmo intenta dividir los datos en un número determinado de grupos (o "clusters") basándose en las similitudes entre ellos. Cada cluster se define por un "centroide", que es el punto medio de todos los puntos de datos en ese cluster. Los puntos de datos se asignan al cluster cuyo centroide esté más cerca.

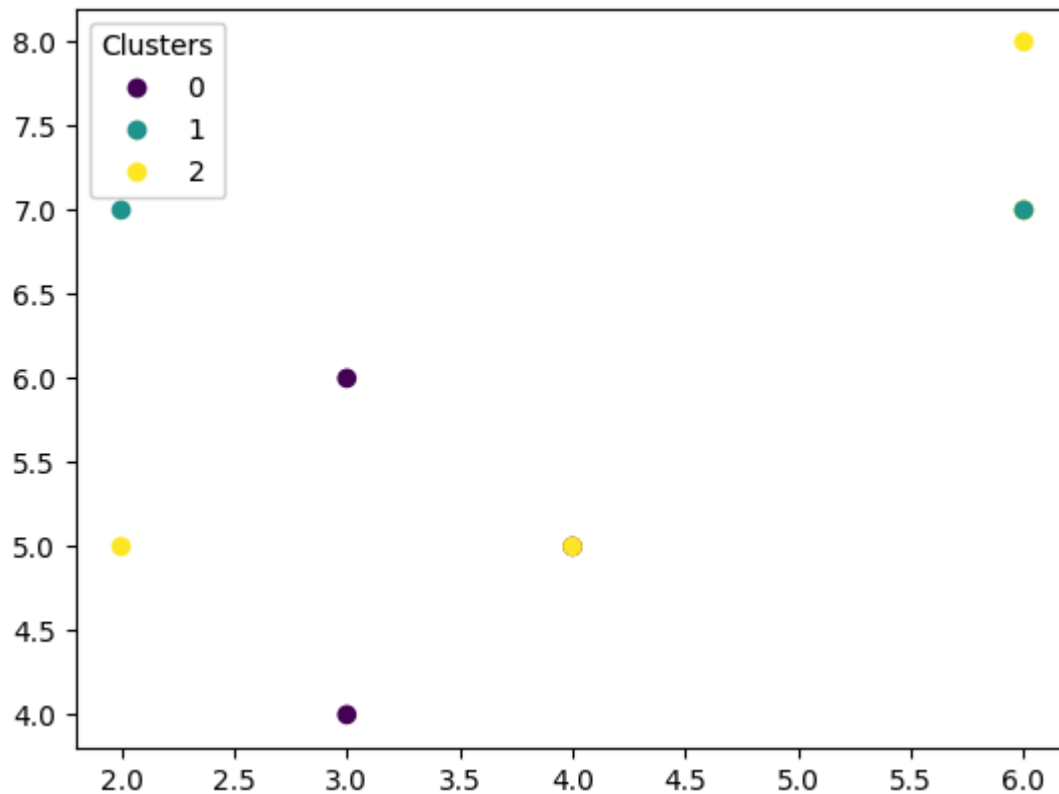
En el código proporcionado, primero creamos una instancia de KMeans con 3 clusters. Luego, ajustamos el modelo a nuestros datos (después de eliminar las columnas 'Hora_Inicio' y 'Hora_Fin'). Finalmente, obtenemos las etiquetas de los clusters para cada punto de datos y las añadimos al DataFrame original.



IBEROAMERICANA

CORPORACIÓN UNIVERSITARIA

```
Estacion_Inicio,Estacion_Fin,Hora_Inicio,Hora_Fin,Duracion,Capacidad_Tren,Pasajeros
4,5,08:00,08:20,20,117,25
6,7,08:15,08:35,20,92,33
4,5,08:30,08:50,20,88,46
3,4,08:45,09:05,20,98,10
2,5,09:00,09:40,40,90,34
6,7,09:15,09:35,20,90,43
3,6,09:30,10:10,40,119,19
2,7,09:45,10:45,60,110,71
6,8,10:00,10:30,30,107,33
6,7,10:15,10:35,20,104,103
```





IBEROAMERICANA

CORPORACIÓN UNIVERSITARIA

```
Estacion_Inicio,Estacion_Fin,Hora_Inicio,Hora_Fin,Duracion,Capacidad_Tren,Pasajeros
1,6,08:00,09:00,60,100,26
7,8,08:15,08:35,20,98,62
2,8,08:30,09:40,70,113,3
7,8,08:45,09:05,20,99,32
2,3,09:00,09:20,20,114,18
5,7,09:15,09:45,30,100,9
3,5,09:30,10:00,30,117,54
7,8,09:45,10:05,20,109,30
4,5,10:00,10:20,20,94,90
5,8,10:15,10:55,40,89,45
```

