

# Agrupamiento de Sesiones de Entrenamiento Deportivo

## Utilizando Mean Shift y Datos Garmin

Sergio Cortés Cepeda

matricula:1731225

Universidad Autónoma de Nuevo León, Facultad de Ciencias Físico Matemáticas

Noviembre 2025

### **Resumen**

El presente artículo analiza datos reales provenientes de entrenamientos de carrera registrados con un dispositivo Garmin. Se aplica el algoritmo de aprendizaje no supervisado Mean Shift con el objetivo de identificar estructuras subyacentes sin requerir la especificación del número de clusters.

## **1. Introducción**

El análisis de datos deportivos se ha fortalecido en los últimos años gracias al uso de dispositivos inteligentes que registran información durante las sesiones de entrenamiento. En este contexto, los algoritmos de aprendizaje no supervisado permiten identificar patrones relevantes sin necesidad de etiquetas previas, facilitando la comprensión del comportamiento a largo plazo.

Entre las técnicas más utilizadas se encuentran DBSCAN y Mean Shift. Mientras DBSCAN detecta densidades de puntos, Mean Shift utiliza estimación de densidad por ventanas para identificar concentraciones naturales en los datos. El propósito de este artículo es aplicar dichos métodos a un conjunto de sesiones de carrera obtenidas con un reloj Garmin, para identificar grupos representativos del comportamiento de entrenamiento.

## 2. Metodología

### 2.1. Datos

El conjunto de datos incluye variables representativas del rendimiento, tales como:

- Distancia recorrida (km)
- Tiempo total (segundos)
- Ritmo medio (seg/km)
- Frecuencia cardiaca media (lpm)
- Cadencia promedio (pasos/min)
- Longitud de zancada (m)
- Desnivel (ascenso/descenso)
- Pasos totales

Inicialmente se evaluó DBSCAN como técnica de densidad. Sin embargo, la dispersión elevada y la heterogeneidad del dataset provocaron que la mayor parte de los puntos fueran clasificados como ruido (-1).

DBSCAN no logró dividir los datos en clusters significativos, por lo que se optó por utilizar otro enfoque.

### 2.2. Algoritmo Mean Shift

Mean Shift estima automáticamente el *bandwidth*, correspondiente al radio de influencia del kernel. Se empleó el método `estimate_bandwidth`

El algoritmo devolvió **cuatro clusters principales**. Este método resultó más adecuado para datos continuos con relaciones no lineales y distribuciones complejas, como los presentes en este caso.

## 2.3. Validación con métricas internas

Se utilizaron tres métricas estándar:

- **Silhouette:** 0.297
- **Calinski-Harabasz:** 6.06
- **Davies-Bouldin:** 0.94

Calinski-Harabasz fue seleccionada como métrica principal debido a su robustez para medir separación y compacidad de clusters.

## 3. Resultados

El algoritmo Mean Shift identificó cuatro grupos:

- **Cluster 0 (Volumen alto):** distancia elevada, cadencia moderada, ritmo estable.
- **Cluster 1 (Alta intensidad):** ritmos muy rápidos, alta FC, sesiones cortas pero exigentes.
- **Cluster 2 (Rodaje suave):** sesiones largas a baja intensidad; FC y ritmo bajos.
- **Cluster 3 (Zancada amplia / técnica trabajada):** mayor longitud de zancada, cadencia media y FC moderada.

Cluster	Distancia (km)	Tiempo (s)	Ritmo (s/km)	Calorías	FC media	FC máx	Cadencia	Zancada (m)	Pasos
0	4.90	1860.2	393.8	404.2	162.4	181.6	163.2	0.94	5154.8
1	5.26	2535.5	487.5	452.0	146.5	185.5	113.5	1.06	5220.0
2	22.58	8565.5	378.5	1835.0	174.5	190.5	171.0	0.92	24545.0
3	8.97	3185.5	349.0	697.0	167.5	187.0	150.0	1.16	8607.0

Cuadro 1: Promedios de variables por cluster utilizando Mean Shift.

Estos clusters fueron interpretados en función de tres dimensiones:

1. **Volumen:** distancia y tiempo total.
2. **Intensidad:** ritmo medio y frecuencia cardiaca.
3. **Técnica:** cadencia y zancada.

## 4. Discusión

Los resultados muestran que Mean Shift permite capturar patrones de entrenamiento que no pueden apreciarse a simple vista. Cada cluster corresponde a un estilo de sesión distinto, lo que permite caracterizar con mayor detalle de mi comportamiento.

La agrupación puede utilizarse para:

- Planear cargas de entrenamiento,
- Identificar tendencias de fatiga,
- Evaluar la técnica de carrera,
- Monitorear progresión fisiológica.

Asimismo, la identificación de cuatro grupos se alinea con las zonas comunes de entrenamiento en corredores recreativos e intermedios, lo cual refuerza la validez del método aplicado.

## 5. Conclusión

El análisis realizado demuestra que Mean Shift es un algoritmo adecuado para identificar estructuras subyacentes en datos deportivos continuos y heterogéneos. El modelo logró agrupar las sesiones en cuatro clusters coherentes con diferentes patrones de volumen, intensidad y técnica.

El uso de métricas internas permitió validar la calidad del agrupamiento, destacando el índice de Calinski-Harabasz como estrategia más estable para este tipo de datos. Los resultados obtenidos pueden servir como base para futuras aplicaciones orientadas al análisis de rendimiento, prevención de lesiones y optimización del entrenamiento.

## Referencias

- [1] DataCamp. (2023). *Mean Shift clustering: A complete guide*. Recuperado de <https://www.datacamp.com/tutorial/mean-shift-clustering>
- [2] GeeksforGeeks. (2023). *DBSCAN Clustering in ML: Density-Based Clustering*. Recuperado de <https://www.geeksforgeeks.org/machine-learning/dbSCAN-clustering-in-ml-density-based-clustering/>