

## DIAPOSITIVA 1.

### ¿De dónde salen los datos?

- Dataset sintético generado con `crear_dataset_actividades_1000()`
- **1000 registros** de señales de sensores (acelerómetro, giroscopio y derivadas)
- **7 actividades humanas:** Caminando, Subiendo/Bajando escaleras, Sentado, De pie, Corriendo, En bicicleta

“Primero generamos y cargamos un dataset sintético de 1000 registros con 7 actividades humanas. Revisamos la estructura con `head`, verificamos que no haya valores nulos y analizamos las estadísticas básicas. Luego vimos la distribución de actividades, que está bastante balanceada, y usamos gráficos (barras, dispersión y matriz de correlación) para entender cómo se agrupan las actividades y qué variables de sensores están más relacionadas.”

## DIAPOSITIVA CLUSTERING

“En la parte no supervisada entrené un modelo de clustering jerárquico. Primero normalicé todas las variables de sensores con `StandardScaler`, porque el algoritmo trabaja con distancias. Luego probé distintos números de clusters y tipos de enlace, y escogí la configuración con mejor silhouette. El modelo final obtiene un silhouette de aproximadamente 0.37 y un ARI de 0.41, lo que indica grupos relativamente compactos y una coincidencia moderada con las actividades reales. Para interpretar los resultados usé una tabla cluster–actividad y una proyección PCA en 2D, donde se ve cómo se separan los clusters.”

## DIAPOSITIVA CONCLUSIONES

### Conclusiones

En el proyecto se probaron varios modelos con datos sintéticos de siete actividades humanas. Entre los supervisados, el SVM con kernel RBF fue el que mejor resultado obtuvo, con la mayor accuracy y F1-score y casi sin errores tras ajustar los hiperparámetros; Random Forest quedó un poco por debajo. El clustering jerárquico ( $\text{silhouette} \approx 0.37$  y  $\text{ARI} \approx 0.41$ ) mostró que los datos tienen estructura y que algunos grupos se parecen a las actividades reales, pero no alcanza la precisión necesaria para usarlo como modelo principal, por lo que SVM se elige como modelo final y el clustering queda como apoyo para análisis y segmentación.

### Recomendaciones

Como mejora futura, se propone repetir el estudio con datos reales de sensores para validar el

rendimiento de SVM, refinar las características utilizadas (más información en frecuencia y en ventanas temporales) y ajustar mejor los modelos de clustering. También es recomendable analizar con más detalle los errores de SVM, sobre todo las confusiones entre subir y bajar escaleras, e incorporar atributos que ayuden a distinguir mejor actividades muy similares.

## **DIAPOSITIVA INTERPRETACION**

### **Interpretación del modelo**

El SVM con kernel RBF logra separar casi perfectamente las 7 actividades humanas. La matriz de confusión muestra solo una confusión entre “Subiendo” y “Bajando escaleras”, lo cual es razonable por lo parecidas que son. Distingue muy bien actividades estáticas de actividades de alta intensidad.

### **Justificación final**

Se elige SVM como modelo final porque obtuvo la mejor accuracy y F1-score, fue estable tras la optimización de hiperparámetros y cometió muy pocos errores. Los modelos de clustering ayudan a entender la estructura de los datos, pero no alcanzan la misma precisión que SVM.