

Reflexión sobre los resultados del análisis no supervisado

El análisis no supervisado, especialmente el análisis de clusters y el Análisis de Componentes Principales (PCA), ofrece valiosos conocimientos sobre las actividades físicas, ayudando a descubrir patrones ocultos en los datos.

Exploración y Preprocesamiento de Datos

El conjunto de datos *Human Activity Recognition with Smartphones Dataset* contiene mediciones de acelerómetros y giroscopios de smartphones mientras los usuarios realizan diversas actividades cotidianas. El objetivo es clasificar las actividades físicas basadas en estas características sensoriales.

El dataset contiene:

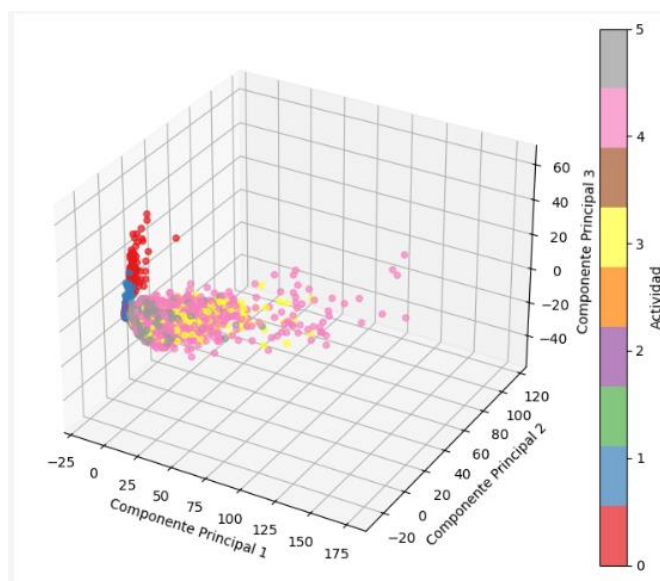
- **Número de variables:** 563 características derivadas de los sensores de acelerómetro y giroscopio en tres ejes (X, Y, Z) para cada tipo de actividad.
- **Número de registros:** 10,299 registros de 30 sujetos diferentes, con mediciones de los sensores para cada instancia de actividad.

Identificación de variables sensoriales:

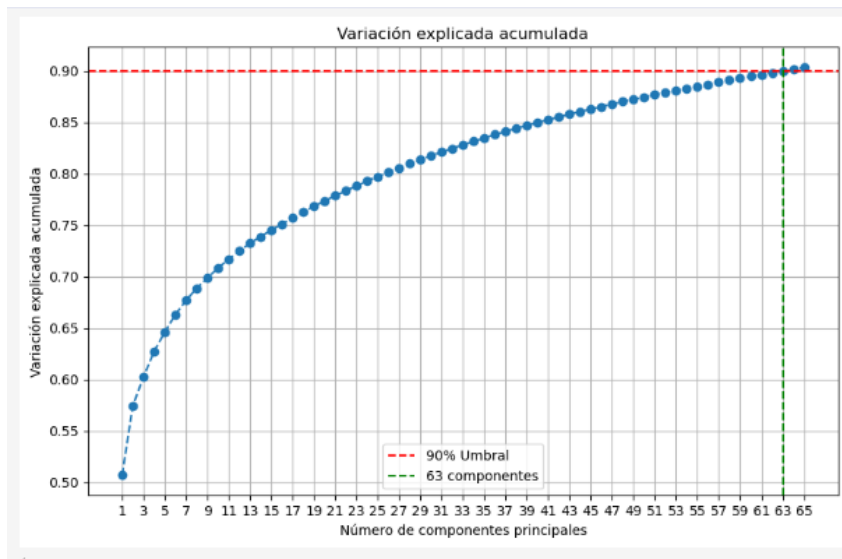
- **Manejo de valores nulos:** No se encontraron valores nulos.
- **Normalización de las lecturas sensoriales:** Se normalizaron las columnas de aceleración para asegurar que las características estén en el mismo rango.

Análisis No Supervisado: PCA

- **Selección de enfoque:** El PCA es apropiado para la reducción de dimensionalidad.

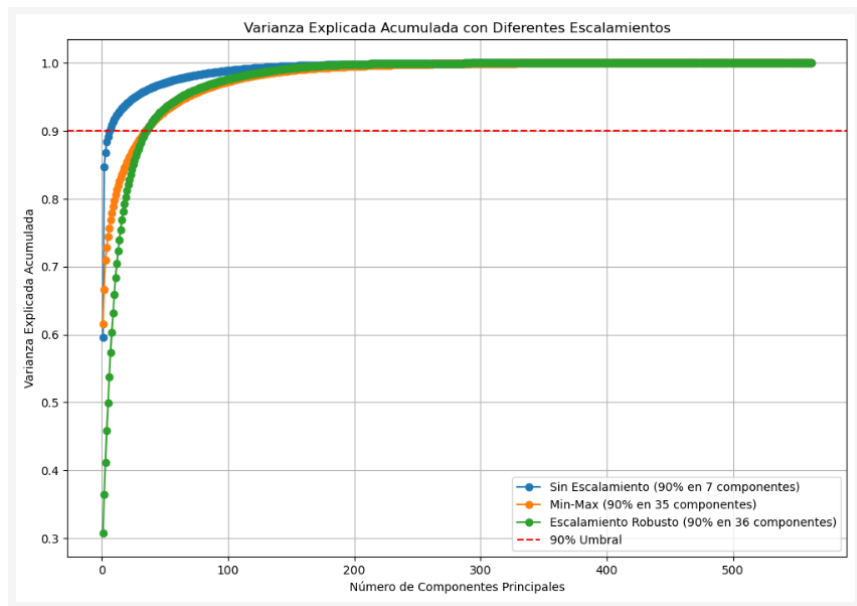


- **PCA (Reducción de dimensionalidad):** Al aplicar PCA, reducimos la dimensionalidad de los datos y visualizamos las relaciones en 2D.



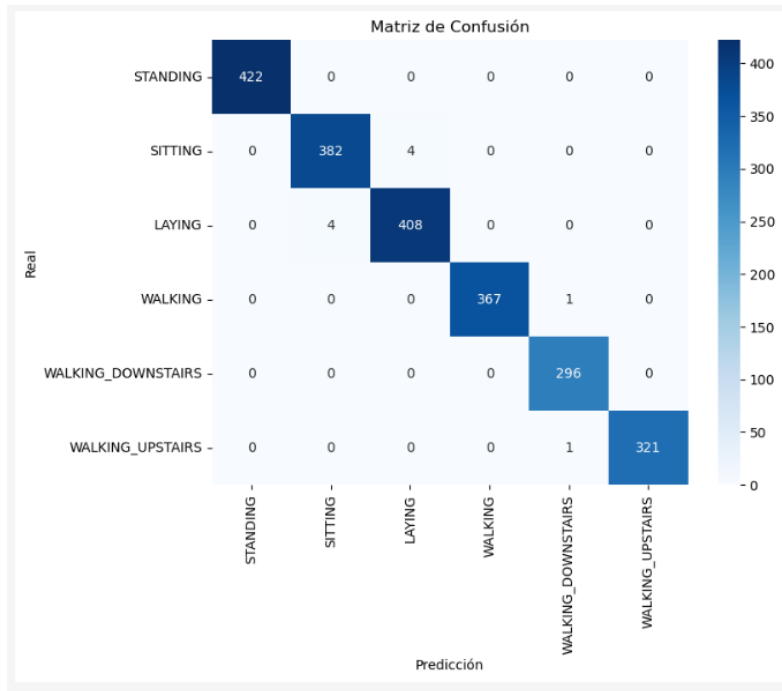
Varianza explicada por los componentes:

- La varianza explicada por cada componente principal indica la cantidad de variabilidad capturada de los datos. Un valor más alto indica que el componente es relevante para describir los datos.
- **Varianza explicada acumulada:** Muestra la suma de las varianzas explicadas por los primeros n componentes principales, ayudando a determinar cuántos componentes son necesarios para describir el modelo de manera eficiente.



Modelado con MLP

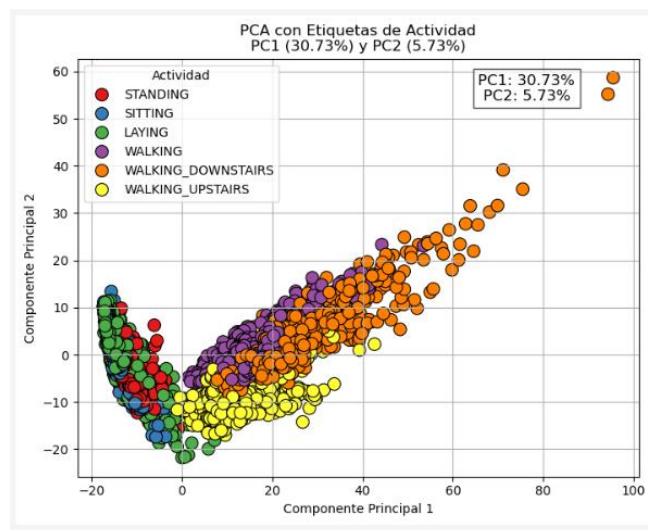
- **Preparación de datos:** Se dividieron los datos en conjuntos de entrenamiento y prueba, asegurando una representación adecuada de las actividades.
- **Evaluación del Modelo:** Se evaluó el modelo utilizando métricas de clasificación, alcanzando un rendimiento excepcional con una precisión del 99.5% en el conjunto de prueba.



Discusión y Análisis

1. Reflexión sobre el análisis no supervisado:

- El uso de PCA permitió reducir significativamente la dimensionalidad (de 563 a 63 dimensiones), manteniendo un alto porcentaje de varianza explicada (90%), lo que mejoró el rendimiento del modelo.



2. Discusión sobre el rendimiento del MLP:

- El MLP mostró un rendimiento excelente con precisión y recall cercanos al 100% en todas las clases. Las métricas F1-score indican un balance ideal entre detectar correctamente las instancias y evitar falsos positivos. Este modelo es altamente efectivo para clasificar actividades físicas.

Precisión en el conjunto de prueba: 0.9954669084315503
Reporte de clasificación:

	precision	recall	f1-score	support
STANDING	1.00	1.00	1.00	422
SITTING	0.99	0.99	0.99	386
LAYING	0.99	0.99	0.99	412
WALKING	1.00	1.00	1.00	368
WALKING_DOWNSTAIRS	0.99	1.00	1.00	296
WALKING_UPSTAIRS	1.00	1.00	1.00	322
accuracy			1.00	2206
macro avg	1.00	1.00	1.00	2206
weighted avg	1.00	1.00	1.00	2206

3. Conclusión:

- Este enfoque es adecuado para predecir actividades físicas utilizando datos sensoriales, especialmente cuando se emplean técnicas de clasificación bien establecidas y reducción de dimensionalidad. Si el rendimiento fuera bajo, sería importante ajustar los hiperparámetros o explorar otros modelos más avanzados.

