

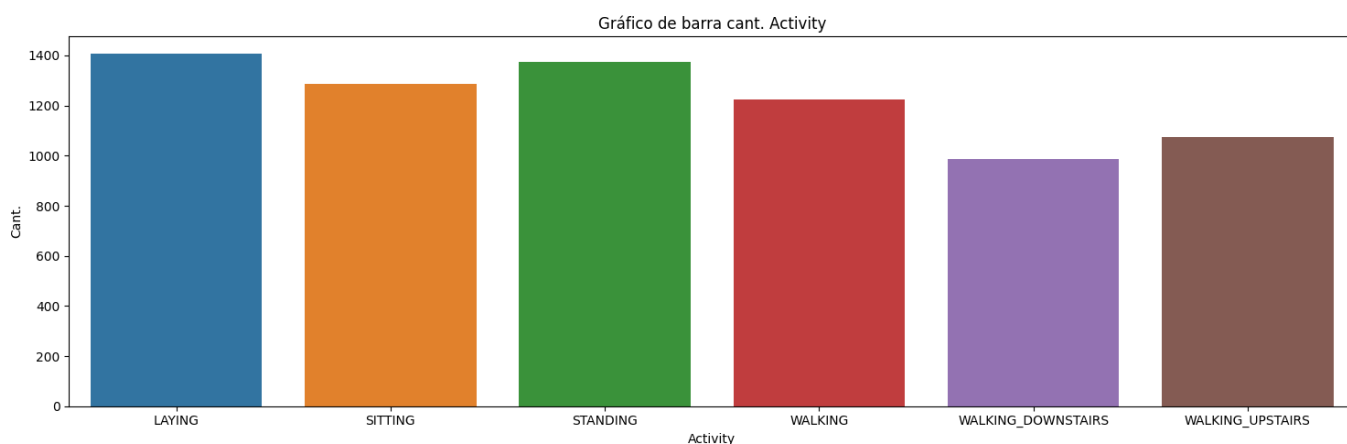
Reporte Examen Cerificación 1 - Human Activity Recognition

Este reporte contiene los resultados de los análisis y evaluación de modelos implementados sobre el dataset Human Activity Recognition, como parte de la resolución del examen.

Exploración y procesamiento de datos

En este paso, se realizó lo siguiente:

- Se descargó el dataset y se analizaron los datos que contiene
- Se verificó que no existían valores nulos que manejar
- No se encontraron valores duplicados
- Se detectó que los datos ya estaban potencialmente normalizados, pero se optó de igual manera probar el PCA con técnicas de escalamiento
- No se detectaron desbalances significativos en los datos que pudieran causar sesgos tanto en el PCA como en el MLP. Esto se puede visualizar abajo.



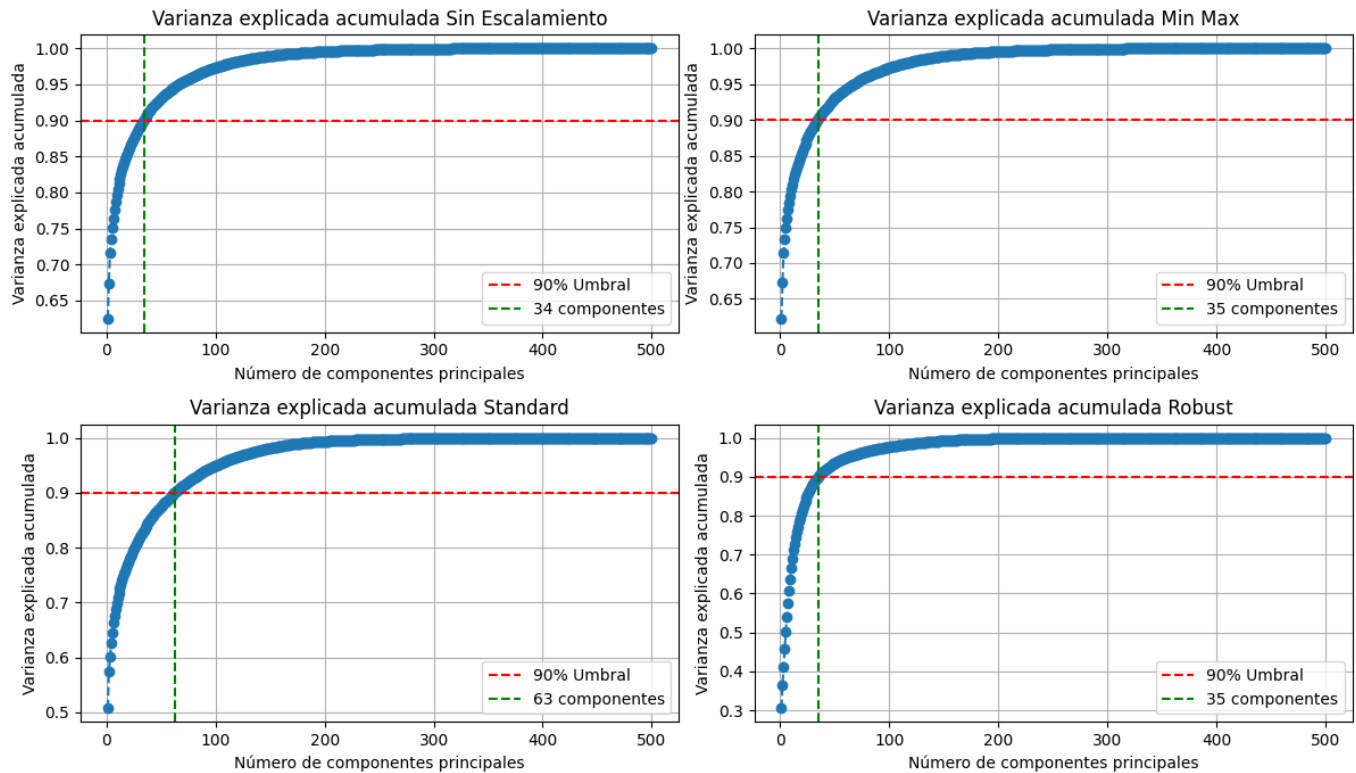
Análisis no supervisado

Se seleccionó la técnica de PCA.

Primeramente, se realizó una búsqueda del número de componentes que permitan mantener el 90% de la varianza. Se utilizaron las siguientes técnicas de escalamiento:

- Sin Escalamiento
- Normalización (MinMaxScaler)
- Estandarización (StandardScaler)
- RobustScaler

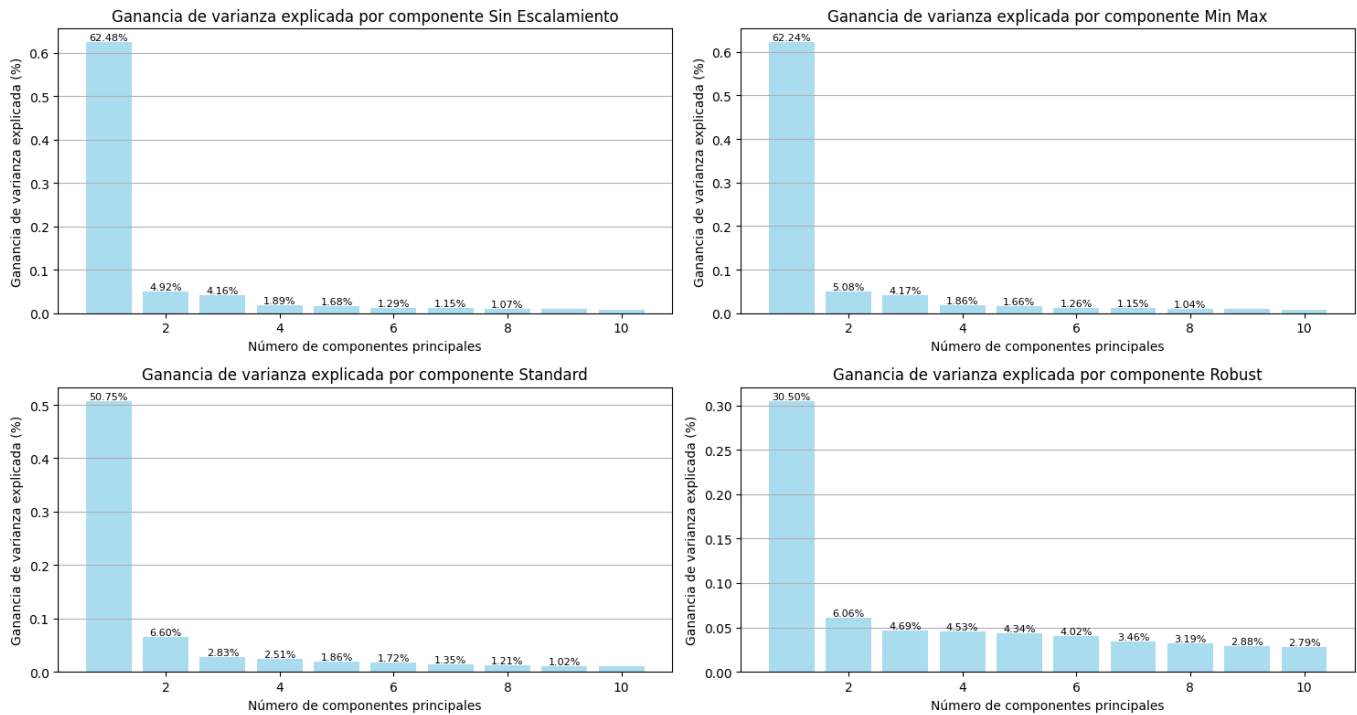
Se pueden visualizar los resultados en el gráfico de abajo.



De este análisis se sacaron las siguientes conclusiones:

- En el caso de no utilizar escalamiento, se puede retener el 90% de la varianza explicada utilizando 34 componentes.
- Usando la normalización como técnica de escalamiento (MinMaxScaler), se logra retener el 90% de la variación usando 35 componentes.
- Usando la estandarización como técnica de escalamiento (StandardScaler), se logra retener el 90% de la varianza usando 63 componentes.
- Finalmente, usando la técnica de escalamiento RobustScaler, se logra retener el 90% de la varianza explicada usando 35 componentes.

También, se graficaron las ganancias de las varianzas explicadas por componente. Este gráfico se encuentra abajo:



De estos gráficos de varianza explicada por componente podemos destacar:

- Para la varianza explicada por componente sin escalamiento:
 - El primer componente es el que más varianza explicada posee, con un 62.48% de varianza explicada acumulada
 - El segundo y tercer componente ya empiezan a proveer menos varianza, solamente alcanzando 4.92% y 4.16% respectivamente
 - Los demas componentes ya empiezan a proveer cada vez menos varianza
- Para la varianza explicada por componente con normalización (MinMaxScaler):
 - Se ven resultados muy similares a los resultados anteriores
 - El primer componente logra capturar 62.24% de la varianza.
 - El segundo y tercer componente acumulan 5.08% y 4.17% de la varianza respectivamente
 - Los demas componentes proveen ya menor varianza
- Para la varianza explicada por componente con estandarización (StandardScaler):
 - Se verifican resultados un poco diferentes a las técnicas anteriores
 - El primer componente sigue manteniendo la mayor cantidad de varianza explicada, aunque el porcentaje disminuyó al 50.72%
 - El segundo componente logra mantener el 6.60% de la varianza explicada
 - En los siguientes componentes ya se empieza a notar una disminución en varianza explicada acumulada.
- Para la varianza explicada por componente con RobustScaler:
 - Se sigue reteniendo la mayor cantidad de la varianza en el primer componente, no obstante el porcentaje es drásticamente menor que en las técnicas anteriores, siendo sólo del 30.5%.
 - El segundo componente logra mantener una varianza explicada del 6.06%
 - Los componentes siguientes logran retener más cantidad de varianza comparado a las otras técnicas de escalamiento, aunque de igual manera los porcentajes son bajos en comparación al componente 1.
- De los 4 gráficos podemos destacar, que existe un componente bastante predominante y otros componentes que retienen menor varianza explicada. Esto nos indica que los valores medidos por los

sensores están en gran parte relacionados, y la técnica de PCA está extrayendo estas relaciones en componentes independientes, lo que resulta en un solo componente acaparando la mayor cantidad de varianza. Esto se podría explicar teniendo en cuenta que los sensores miden variables que están de una u otra manera relacionados. Por ejemplo, la aceleración, velocidad, posición, rotación.