

REPORTE-Reducción de Dimensionalidad con PCA

1. Justificación del algoritmo

Se aplicó el método **PCA (Análisis de Componentes Principales)** porque es una técnica eficaz para **reducir la dimensionalidad** de un conjunto de datos, preservando la mayor parte de la información original (varianza).

En este caso, el dataset contiene múltiples variables numéricas como *ingresos*, *gastos_comunes*, *pago_coche*, *ahorros*, *vivienda*, etc. Reducir la dimensionalidad permite simplificar el análisis y facilita la visualización de patrones, eliminando redundancias y correlaciones fuertes entre variables.

2. Descripción del diseño del modelo

1. **Selección de variables numéricas:** Se incluyeron únicamente las columnas de tipo numérico para aplicar PCA de forma correcta.
2. **Estandarización:** Se aplicó `StandardScaler` para escalar todas las variables a media 0 y desviación estándar 1, evitando que las variables con valores más grandes dominan el cálculo.
3. **Aplicación de PCA:** Se configuró `n_components=0.95`, lo que significa que se seleccionan tantas componentes principales como sean necesarias para explicar al menos el **95% de la varianza total**.
4. **Visualización:** Se graficó la proyección de los datos en las dos primeras componentes principales para identificar posibles agrupaciones o patrones.
5. **Guardado del modelo:** Tanto el objeto PCA como el escalador fueron guardados con `joblib` para poder reutilizarlos en el futuro.

4. Gráfica personalizada e interpretación de resultados

- **Interpretación:** La dispersión de puntos en la gráfica 2D muestra cómo se distribuyen los registros según las dos componentes principales.
- Las variables *ingresos*, *ahorros* y *vivienda* probablemente tienen alta carga en la primera componente, asociada al nivel económico general.
- Las variables de *gastos* y *pago_coche* podrían tener mayor peso en la segunda componente, relacionada con los patrones de consumo.
- La gráfica permite identificar perfiles financieros distintos entre los registros, lo que podría servir como base para segmentación o modelos predictivos.