

Actividad

Cada grupo debe trabajar en su proyecto final incorporando las siguientes etapas:

1. Limpieza avanzada del dataset

- Detectar y eliminar duplicados, valores nulos y outliers.
- Justificar las transformaciones aplicadas.
- Mostrar visualmente el “antes y después” de limpiar los datos.

2. Balanceo de clases

- Si su dataset es de clasificación, aplicar **SMOTE** u otra técnica de oversampling o undersampling.
- Mostrar cómo cambia la proporción de clases con un gráfico.
- Explicar cómo el balanceo mejora el entrenamiento.

3. Selección de atributos

- Usar técnicas como *SelectKBest*, *Feature Importance* o *correlaciones*.
- Mostrar qué variables aportan más valor y cuáles pueden eliminarse.
- Explicar por qué mantener solo las variables relevantes.

4. Reducción de dimensionalidad

- Aplicar **PCA** (Análisis de Componentes Principales) para reducir variables.
- Visualizar los datos transformados (por ejemplo, en 2D).
- Explicar en qué mejora la interpretabilidad o velocidad del modelo.

5. Visualización y comunicación de resultados

- Crear gráficos y conclusiones que expliquen el “viaje del dataset” desde su estado crudo hasta el modelo optimizado.
- Presentar el proceso y los resultados en lenguaje claro y visual.

Deben realizar una **presentación técnica de máximo 20 minutos** con soporte visual (gráficos, ejemplos, comparaciones, conclusiones), explicar con código.

Todos los integrantes deben participar activamente en la exposición.

Esta actividad equivale al 5% de la nota 2.