Leonel Moises Chauca Maydana

**Objetivos Claros para el Proyecto**

1. **Objetivo General:**
   * Desarrollar y evaluar un modelo de clasificación de géneros de videojuegos utilizando técnicas de aprendizaje supervisado y no supervisado, con un enfoque en el preprocesamiento de datos, balanceo de clases, y reducción de dimensionalidad para mejorar el rendimiento y la precisión del modelo.
2. **Objetivos Específicos:**
   1. **Preprocesamiento y Limpieza de Datos:**
      * Realizar el preprocesamiento de datos, incluyendo la gestión de valores faltantes en la columna genre, utilizando técnicas como el reemplazo con valores predeterminados y la creación de una columna de géneros agrupados.
      * Aplicar técnicas de codificación como **One-Hot Encoding** para convertir variables categóricas en características numéricas, y **normalización** para asegurar que todas las características estén en una escala adecuada para el modelado.
   2. **Balanceo de Clases:**
      * Implementar técnicas de **oversampling** para balancear las clases minoritarias, replicando instancias mediante el uso de la librería resample de Python, de forma que todas las clases tengan una representación equitativa en el conjunto de datos de entrenamiento.
      * Evaluar el impacto de estas técnicas en el rendimiento del modelo, verificando la distribución de clases antes y después del balanceo.
   3. **Entrenamiento y Evaluación del Modelo Supervisado:**
      * Utilizar un **Random Forest** para la clasificación de los géneros de videojuegos y evaluar su rendimiento utilizando métricas como **precisión**, **recall**, **F1-score**, y la **matriz de confusión**.
      * Realizar múltiples ejecuciones con **splits** de entrenamiento/prueba (80/20 y 50/50) para medir la consistencia y confiabilidad del modelo.
   4. **Optimización mediante Reducción de Dimensionalidad (PCA):**
      * Aplicar **PCA** (Análisis de Componentes Principales) para reducir la dimensionalidad de los datos, probando distintas configuraciones con 12, 10, 9, 5 y 3 componentes principales.
      * Evaluar el impacto de la reducción de dimensiones en la **varianza explicada** y en el rendimiento del modelo de clasificación, comparando los resultados obtenidos antes y después de aplicar PCA.
   5. **Evaluación No Supervisada:**
      * Realizar un **análisis no supervisado** utilizando técnicas como PCA y **clustering** para explorar patrones subyacentes en los datos sin utilizar las etiquetas de clase.
      * Analizar cómo los datos se agrupan naturalmente y evaluar la calidad de esos grupos, para identificar posibles patrones o relaciones adicionales entre las características.