

Descripción de la Actividad: Aplicación de Machine Learning a Problemas Reales

Objetivo:

Los estudiantes trabajarán en grupos para buscar conjuntos de datos relevantes y desarrollar una presentación que explique un problema específico en alguna de las siguientes áreas:

- **Transición energética justa:** Democratización de la generación y consumo energético, desarrollo de comunidades energéticas, impulso a energías limpias (hidrógeno verde, eólica, solar, etc.) y minerales estratégicos.
- **Ciencia, tecnología e innovación para la transformación productiva:** Resolución de desafíos sociales, económicos y ambientales mediante la innovación y el uso de datos.

El propósito es que identifiquen a los interesados (stakeholders) y propongan cómo un modelo de **machine learning supervisado** podría abordar el problema, aplicando algoritmos adecuados, especialmente de **predicción**.

Pasos de la Actividad

1. Búsqueda de Conjuntos de Datos

- Trabaja en grupo y busca conjuntos de datos en plataformas recomendadas como:
 - [Kaggle](#)
 - [UCI Machine Learning Repository](#)
 - Bases de datos gubernamentales o académicas.
- Asegúrate de que los datos sean **etiquetados**, es decir, que tengan una variable objetivo o de salida que permita aplicar modelos de **aprendizaje supervisado**.

Ejemplos de conjuntos de datos relevantes:

- Datos sobre consumo energético y fuentes de energía limpia.
 - Datos sobre contaminación ambiental y su impacto en comunidades.
 - Datos sobre producción y uso de minerales estratégicos.
 - Datos sobre innovación tecnológica y su impacto en sectores productivos.
-

2. Identificación del Problema

- Analiza el conjunto de datos elegido y define un **problema específico** que pueda abordarse con machine learning.
- Identifica los **stakeholders** afectados, como comunidades, empresas, gobiernos, investigadores, entre otros.

Ejemplo de problemas que se pueden abordar:

- **Predicción del consumo energético en comunidades rurales** para optimizar el uso de energías renovables.
 - **Clasificación de tipos de contaminación en zonas industriales** para tomar medidas preventivas.
 - **Predicción de la demanda de minerales estratégicos** para optimizar la producción.
-

3. Propuesta de Solución con Machine Learning

- Explica cómo un modelo de **machine learning supervisado** podría resolver el problema.
- Discute qué tipo de modelo sería adecuado:
 - **Regresión**: si el problema implica predecir un valor numérico (por ejemplo, consumo energético futuro).
 - **Clasificación**: si el problema implica categorizar datos (por ejemplo, clasificar tipos de contaminación en diferentes áreas).

Ejemplo de modelos a utilizar:

- **Regresión lineal o árboles de decisión** para predecir el consumo energético.

- **Random Forest o SVM** para clasificar tipos de contaminación.
-

4. Desarrollo de la Presentación

Cada grupo debe preparar una presentación clara y concisa que incluya:

- ✓ **Descripción del problema y su relevancia.**
 - ✓ **Identificación de los interesados y cómo se ven afectados.**
 - ✓ **Análisis del conjunto de datos y variables más importantes.**
 - ✓ **Propuesta de solución con machine learning**, indicando el tipo de modelo y enfoque general.
 - ✓ **Visualizaciones de datos** para respaldar la propuesta (gráficos, mapas, etc.).
-

5. Presentación y Discusión

- Cada grupo presentará sus hallazgos y propuestas al resto de la clase.
 - Se abrirá una discusión para analizar los enfoques presentados, recibir retroalimentación y sugerencias.
-

Conjuntos de Datos y Librerías Recomendadas

Conjuntos de datos en Kaggle:

- [World Energy Consumption](#) (Consumo energético global).
- [Global Power Plant Database](#) (Plantas de energía en el mundo).
- [Air Quality Data](#) (Datos sobre contaminación del aire).

Conjuntos de datos en UCI Repository:

- [Energy Efficiency Dataset](#) (Eficiencia energética de edificios).

Librerías recomendadas:

- **pandas** → para manipulación de datos.
- **numpy** → para operaciones matemáticas.

- `scikit-learn` → para modelos de machine learning supervisado.
- `matplotlib` y `seaborn` → para visualización de datos.