



TALENTO
TECH

▶▶▶▶

Misión 1

INTELIGENCIA ARTIFICIAL

Explorador



▶▶▶▶

Reto misión 1



DESARROLLO DEL PROYECTO

En esta práctica, los estudiantes deben crear un video en el que presenten un proyecto de machine learning desde su inicio hasta el análisis de datos. Primero, deben ofrecer una breve introducción teórica sobre las fases del ciclo de vida de un proyecto de machine learning, discutiendo técnicas para identificar problemas y la importancia de los stakeholders. Luego, deben seleccionar un contexto relevante que esté en el marco de las siguientes 2 líneas:

- Transición energética justa, democratización de la generación y consumo energético, desarrollo de comunidades energéticas, impulso a energías limpias (hidrógeno verde, eólica, solar, entre otras) y minerales estratégicos.
- Ciencia, tecnología e innovación para la transformación productiva y la resolución de desafíos sociales, económicos y ambientales, contribuyendo a la construcción de una sociedad del conocimiento.

A continuación, mostrarán cómo realizar un análisis de datos, incluyendo la carga y exploración del conjunto de datos, identificación de tipos de datos, manejo de datos faltantes, y normalización. Finalmente, deben reflexionar sobre el proceso seguido, los desafíos encontrados y las lecciones aprendidas.

Nota: Se recomienda que el dataset sea acorde para un problema de clasificación o regresión con datos multivariados. (Esto es importante definirlo con miras a la práctica final)

COMPOSICIÓN DEL PROYECTO DE VIDEO

Introducción Teórica

- Repasar las diferentes fases del ciclo de vida de un proyecto de machine learning.
- Discutir las técnicas y metodologías utilizadas para identificar problemas en machine learning.
- Destacar la importancia de identificar a los stakeholders (interesados) y su rol en el proyecto.

Selección de Contexto y Problema

- Revisar varios ejemplos de contextos regionales tales como:
 - Transición energética justa, democratización de la generación y el consumo energético, desarrollo de comunidades energéticas, impulso a las energías limpias, (hidrógeno verde, eólica, solar, entre otras) y minerales estratégicos.
 - Ciencia, tecnología e innovación para la transformación productiva y la resolución de desafíos sociales, económicos y ambientales del país; y la construcción de una sociedad del conocimiento.
- Discutir por qué este problema es relevante.
- Identificar a los interesados

Análisis de Datos Básicos

Carga y Exploración del Conjunto de Datos:

- Mostrar cómo cargar el conjunto de datos en un DataFrame utilizando pandas.
- Revisar la cantidad de columnas y filas para tener una idea general del tamaño del conjunto de datos.

Tipos de Datos:

- Identificar los tipos de datos de cada columna (numéricos, categóricos, booleanos, etc.).
- Discutir la importancia de los tipos de datos en el análisis y modelado.

Cantidad de Datos Faltantes:

- Detectar la cantidad de datos faltantes en cada columna.
- Discutir sobre el impacto que tiene los datos faltantes en el problema en cuestión en fases más avanzadas del desarrollo del modelo de machine learning

Verificación de Reglas de Rango y Dominio:

- Identificar y marcar los datos basura, como valores erróneos o irrelevantes.
- Revisar valores fuera de un rango esperado o texto en columnas numéricas.
- Discutir estrategias para limpiar los datos basura.

Dimensiones de la Calidad de Datos:

- Evaluar la completitud, consistencia y exactitud de los datos.

Detección y Tratamiento de Datos Ausentes:

- Mostrar cómo manejar los datos faltantes mediante el descarte o técnicas de imputación (media, mediana, regresión, Hot deck).

Normalización de Datos:

- Aplicar técnicas de Min-Max Scaling y escalado robusto para normalizar los datos y mejorar el rendimiento del modelo.

Análisis Univariable:

- Validar la distribución normal de las variables.
- Calcular estadísticas descriptivas como media, mediana, moda y desviación estándar.

Análisis Bivariado:

- Calcular y visualizar la correlación entre variables para explorar la relación entre dos variables.

Análisis Multivariado:

- Realizar un Análisis de Componentes Principales (PCA) para reducir la dimensionalidad.
- Crear visualizaciones complejas para entender mejor la interacción entre múltiples variables.

Reflexión y conclusión

- Reflexionar sobre el proceso seguido y los desafíos encontrados.