**GUION INTRODUCCIÓN COMPONENTE FORMATIVO 03**

Título del video: **Visualización, partición y automatización en el procesamiento de datos**

El ciclo de vida del **aprendizaje automático** no comienza con el modelo, sino con los **datos**. Este proceso abarca desde la **recolección** inicial hasta la **evaluación y automatización del flujo de trabajo**, permitiendo el desarrollo de soluciones más precisas y confiables.

Primero, se realiza una **limpieza y transformación de los datos**. Esto incluye el tratamiento de **valores atípicos, la normalización y la conversión de formatos**, tareas fundamentales para asegurar que los algoritmos funcionen correctamente. Posteriormente, los datos se **dividen en conjuntos de entrenamiento y prueba**, y se **seleccionan modelos** que serán evaluados con métricas como la **precisión** o el **error cuadrático medio**.

Una de las prácticas clave en esta etapa es la automatización mediante ***pipelines***, como los que ofrece **Scikit-learn**. Estas estructuras permiten **encadenar procesos de forma ordenada y reproducible**, garantizando que todos los datos pasen por las mismas **etapas de transformación**.

Asimismo, herramientas como **MLflow** permiten **monitorear experimentos, registrar modelos, evaluar su desempeño y facilitar la toma de decisiones**. De manera complementaria, aplicar buenas prácticas de **gobernanza** de datos asegura **calidad, trazabilidad y cumplimiento** normativo en todo el proceso.

Comprender y dominar este ciclo es esencial para desarrollar modelos eficaces, reducir errores y optimizar el uso de recursos.

TOTAL PALABRAS: 219