

# Tuberías en Machine Learning

---

## ¿Por qué usar tuberías?

- Organización del flujo de trabajo: permite agrupar y visualizar todos los pasos necesarios desde el preprocesamiento hasta el entrenamiento del modelo.
- Reutilización de código: el mismo pipeline puede aplicarse a nuevos datos sin necesidad de reescribir los pasos.
- Reducción de errores: evita aplicar transformaciones diferentes en los conjuntos de entrenamiento y prueba.
- Facilidad de validación cruzada: permite evaluar el rendimiento del modelo con mayor robustez.
- Mejor mantenimiento y escalabilidad: los procesos se estandarizan, facilitando la incorporación de mejoras.

## Componentes técnicos de una tubería

- Preprocesamiento de datos: imputación de valores faltantes, normalización o estandarización de variables, codificación de variables categóricas.
- Transformación o selección de características: extracción y selección de variables relevantes.
- Modelo de aprendizaje: entrenamiento y predicción mediante algoritmos supervisados o no supervisados.
- Evaluación y validación: integración con validación cruzada, grid search u otros esquemas.
- Predicción y despliegue: uso del pipeline entrenado para hacer inferencias sobre nuevos datos.

## Especificaciones técnicas

- Modularidad: permite dividir el proceso en módulos reutilizables.
- Compatibilidad: se integra con bibliotecas como scikit-learn, TensorFlow, Spark ML, entre otras.
- Persistencia: las tuberías pueden guardarse y reutilizarse.
- Trazabilidad: cada paso es rastreable, facilitando el depurado.
- Paralelización: algunas implementaciones permiten procesar en paralelo para mayor eficiencia.

## Ventajas

- Mayor claridad y legibilidad del código.
- Reducción de errores humanos.
- Consistencia entre entrenamiento y producción.
- Facilidad para experimentar con nuevos algoritmos.
- Mejora en rendimiento y escalabilidad.

### Desventajas o limitaciones

- Curva de aprendizaje inicial.
- Dificultad para representar pasos complejos.
- Menor flexibilidad en intervenciones manuales.
- Rendimiento dependiente de la optimización de pasos.

### Conclusión

- Las tuberías son esenciales para estructurar eficientemente el proceso de entrenamiento y evaluación en Machine Learning. Automatizan tareas repetitivas, garantizan coherencia de datos y mejoran la trazabilidad, siendo clave en entornos de producción o experimentación avanzada.