

DOCUMENTACIÓN: PERSISTENCIA DE MODELOS EN PYTHON

Objetivo: Investigar y probar métodos para guardar y recuperar modelos de regresión lineal en Python, asegurando su correcta funcionalidad tras ser cargados.

1. Métodos investigados:

- pickle (módulo estándar de Python)
- joblib (biblioteca optimizada para modelos grandes)

2. Comparativa de ventajas y desventajas:

Pickle:

- Biblioteca estándar de Python.
- Adecuado para modelos pequeños o prototipos.
- Más lento con grandes volúmenes de datos.
- Archivos más grandes.
- No requiere instalación.
- Fácil de usar (dump y load).

Joblib:

- Optimizado para modelos grandes y arrays numéricos.
- Más rápido que pickle.
- Archivos más pequeños (compresión eficiente).
- Requiere instalación: pip install joblib.
- Muy usado con scikit-learn.
- Sintaxis casi idéntica a pickle

3. Conclusión:

Ambos métodos funcionan correctamente para guardar y recuperar modelos. Joblib es más recomendable para proyectos de Machine Learning por su eficiencia con modelos grandes. Pickle es útil para pruebas simples o entornos sin dependencias externas.

4. Proceso de guardado y recuperación de un modelo:

- Crear y entrenar el modelo con los datos de entrada.
- Guardar el modelo en un archivo usando pickle o joblib según corresponda.
- Asegurarse de que el archivo se encuentra en un directorio accesible para los usuarios.
- Para recuperar el modelo, cargarlo desde el archivo usando la misma biblioteca.
- Verificar que el modelo cargado puede realizar predicciones correctamente.

5. Bibliotecas utilizadas:

- Numpy
- Scikit_learn
- Pickle
- Joblib

