La generación de código para Machine Learning es una parte vasta y en la que se puede profundizar increíblemente, sin embargo, en el gran esquema de las cosas implica una pequeña parte del ciclo de vida entero de un proyecto.

En las siguientes entregas de esta saga se interiorizará sobre el resto de etapas, donde aprenderás a profesionalizarlas.



Ya tienes tus códigos, tus configuraciones y pesos, pero, ¿Cómo los haces accesibles al usuario final?

**Ejemplos de producción**

Puedes desplegar tus modelos en diferentes dispositivos según tu necesidad.

Si tu proyecto va a ser de consumo masificado, entonces la opción natural será desplegarlo en la nube, donde Google Cloud, Azure, AWS u Oracle Cloud podrán ayudarte. Esta ventaja es especialmente útil si debes escalar tu modelo a mayores capacidades sin necesidad de adquirir un equipo propio.

Si necesitas hacer inferencias en vivo entonces podrías optar por equipo IoT, donde dispositivos como la Raspberry Pi o el Jatson Nanon te ofrecerán una capacidad de cómputo decente para tareas en tiempo real.

Si tienes los recursos necesarios o el proyecto no es tan robusto, puedes correr tus modelos de manera local, donde tus equipos se encargarán de las inferencias.

Un caso final (y una extensión a los últimos 2 casos) sería el de usar un USB Accelerator, hardware con alta capacidad de cómputo que procesa las inferencias con alta facilidad.

Puedes concentrar los recursos de predicción sobre este hardware y dejar descansar al resto del equipo.

**Contribución creada por** Sebastián Franco Gómez.