

Tecnologías para MLOps

Docker

Docker es una plataforma open source que se utiliza para automatizar el despliegue a producción y manejar aplicaciones dentro de contenedores aislados. Es bastante ligero en comparación de las máquinas virtuales tradicionales. Docker es importante para MLOps porque:

1. Ayuda a mejorar la colaboración, reduciendo dependencias mediante el empaquetamiento de las herramientas necesarias y librerías.
2. Reduce la confusión entre los equipos de desarrollo y producción, por medio de un ambiente consistente y de fácil reproducibilidad.
3. Simplifica la escalabilidad de las aplicaciones de ML ya que cuenta con un modelo consistente de despliegue. Los contenedores pueden replicarse fácilmente y distribuirse a través de instancias permitiendo una utilización eficiente de recursos.

Algunas alternativas a Docker son:

1. Podman: simple y seguro, ofrece una interfaz de línea de comandos compatible con Docker. Permite administrar contenedores y pods. Es seguro en entornos compartidos.
2. Rkt (Rocket): desarrollado con los estándares de contenedores de la Open Container Initiative (OCI).
3. LXC (Linux containers): permite crear y administrar contenedores en sistemas Linux. Permite un mayor control sobre los contenedores.

PyCaret

PyCaret es una librería de machine learning de código abierto, low-code que se utiliza para construir pipelines de extremo a extremo de machine learning en Python. Esta librería es conocida por su facilidad de uso, simplicidad y la habilidad de construir y desplegar rápida y eficientemente pipelines de machine learning.

PyCaret se utiliza en varias etapas del ciclo de vida de ML y MLOps:

1. Preparación de datos: proporciona funciones para cargar, explorar y preprocesar datos de manera eficiente.
2. Selección de modelos y ajuste de hiperparámetros: permite comparar varios algoritmos de ML y realizar la optimización de hiperparámetros.
3. Generación de informes y visualización: crea informes detallados y visualizaciones que ayudan en la evaluación y comprensión de los modelos.
4. Despliegue de modelos: facilita la exportación de modelos entrenados para su despliegue en producción.

Dentro de las alternativas podemos encontrar:

1. Scikit-learn: biblioteca de aprendizaje automático de Python con una gran cantidad de algoritmos de aprendizaje supervisado y no supervisado, herramientas para la selección de modelos y evaluación de rendimiento.
2. TPOT (tree based pipeline optimización tool): biblioteca de AutoML que usa algoritmos genéticos para optimizar pipelines de ML, incluyendo la selección de modelos y ajuste de hiperparámetros.
3. Kubeflow: proyecto de código abierto para MLOps en Kubernetes, utilizado para gestionar pipelines de entrenamiento y despliegue de modelos en entornos de producción.

MLFlow

MLFlow es una plataforma de código abierto que se utiliza para manejar el ciclo de vida de machine learning, incluyendo la experimentación, reproducibilidad, despliegue y registro de modelo central. Ofrece 4 componentes:

1. Tracking: graba y busca experimentos: código, data, configuraciones, resultados.
2. Projects: empaqueta código de data science en un formato que permite reproducir las corridas en cualquier plataforma.
3. Models: despliega modelos de machine learning en diversos ambientes.
4. Registry: Almacena, anota, descubre y maneja modelos en un repositorio central.

Como alternativas a MLFlow tenemos:

1. Kubeflow: plataforma de código abierto para orquestar y gestionar el ciclo de vida completo de los proyectos de ML en kubernetes. Cuenta con herramientas de creación de pipelines de entrenamiento, el despliegue y la monitorización de modelos en producción.
2. DVC (data versión control): herramienta de control de versiones diseñada para el seguimiento y gestión de datos de ML.
3. TFX (TensorFlow extended): plataforma de código abierto de Google centrada en la producción de modelos de ML escalables. Proporciona componentes para la construcción de pipelines de MLOps y herramientas para la validación y evaluación de modelos de producción.

DagsHub

DagsHub es un GitHub para machine learning. Es una plataforma para data scientists y desarrolladores de machine learning para versionar su data, modelos, experimentos y código. Provee una experiencia similar a la de GitHub para machine learning, permitiendo que los equipos compartan, revisen y reusen el trabajo creado de manera rápida. Debido a que esta construido en herramientas y formatos populares de código abierto es muy fácil integrar con herramientas conocidas.

DVC

Data Version Control o DVC es una herramienta experimental para proyectos de ML. Esta basado en Git y su objetivo principal es codificar data, modelos y pipelines a través de la línea de comando. DVC reemplaza archivos grandes como los datasets o modelos de ML con pequeños metaarchivos que apuntan a la data original. DVC tambien puede manejar los pipelines de los proyectos para hacer que los experimentos sean reproducibles por todos los miembros del equipo.

Además de DVC, estas herramientas son algunas alternativas:

1. Git-LFS (git large file storage) : es una extension de Git que permite gestionar archivos grandes en repositorios de Git. No cuenta con todas las características de DVC pero es una alternativa simple para gestionar datos grandes junto con un control de versiones.

2. CML (continuous machine learning): Herramienta de GitHub que permite automatizar flujos de trabajo de ML directamente en repositorios de GitHub. Se puede usar con Git y GitLFS para gestionar datos y modelos en un entorno colaborativo de ML.
3. Pachyderm: plataforma de datos y control de versiones para ML. Ofrece una manera de versionar datos y ejecutar pipelines de datos para facilitar la construcción y gestión de flujos de trabajo de ML.

CookieCutter

Cookiecutter es un paquete de Python, de fácil instalación, que permite crear y usar plantillas para microservicios y proyectos de desarrollo. Es una herramienta de línea de comando que no requiere conocimiento de Python para poder utilizarse.

Cookiecutter se puede usar para definir una estructura de proyecto estándar que ayude para la consistencia en la organización de proyectos. Con cookiecutter se puede definir plantillas de código que tengan estructuras de directorios, scripts de inicio, configuraciones de entorno entre otras para proyectos de ML. También ayuda a garantizar la reproducibilidad en proyectos de ML.

CookieCutter tiene las siguientes alternativas:

1. Yeoman: generador de proyectos web que permite crear estructura de proyectos a partir de generadores predefinidos. Su enfoque es para proyectos web, pero se puede adaptar a proyectos de ML.
2. Hygen: herramienta de generación de código que se utiliza para crear y mantener plantillas de proyectos. Ofrece una forma simple de generar archivos, código y estructuras de proyectos a partir de plantillas personalizables.
3. Custom scripts, makefiles: scripts personalizados o makefiles para crear y configurar estructuras de proyectos de ML.

Flask

Flask es un framework de micro web basado en Python. Con Flask podemos crear web apps que se pueden utilizar para consumir el modelo creado. Un usuario puede ingresar valores en un formulario y después de enviarlo, el modelo va a predecir el resultado.

Flask se usa en MLOps para crear aplicaciones web con APIs para servir modelos, con paneles de control y visualización de datos, aplicaciones de gestión de proyectos y flujos de trabajo; puede estar detrás de aplicaciones web que ofrecen control de versiones y seguimiento de modelos y datos en MLOPs; se utiliza para agregar autenticación y autorización a las aplicaciones web de MLOps garantizando que solo usuarios autorizados puedan acceder a la aplicación.

Además de Flask, podemos utilizar:

1. FastAPI: framework moderno y rápido para crear aplicaciones web en Python. Se usa como alternativa rápida y eficiente de Flask, cuando se busca un rendimiento óptimo en aplicaciones de MLOps.
2. Django: Mas completo que Flask, proporciona una amplia gama de características como una capa de administración de datos y autenticación.
3. Starlette: framework asíncrono en Python que se usa para crear aplicaciones web rápidas y escalables.