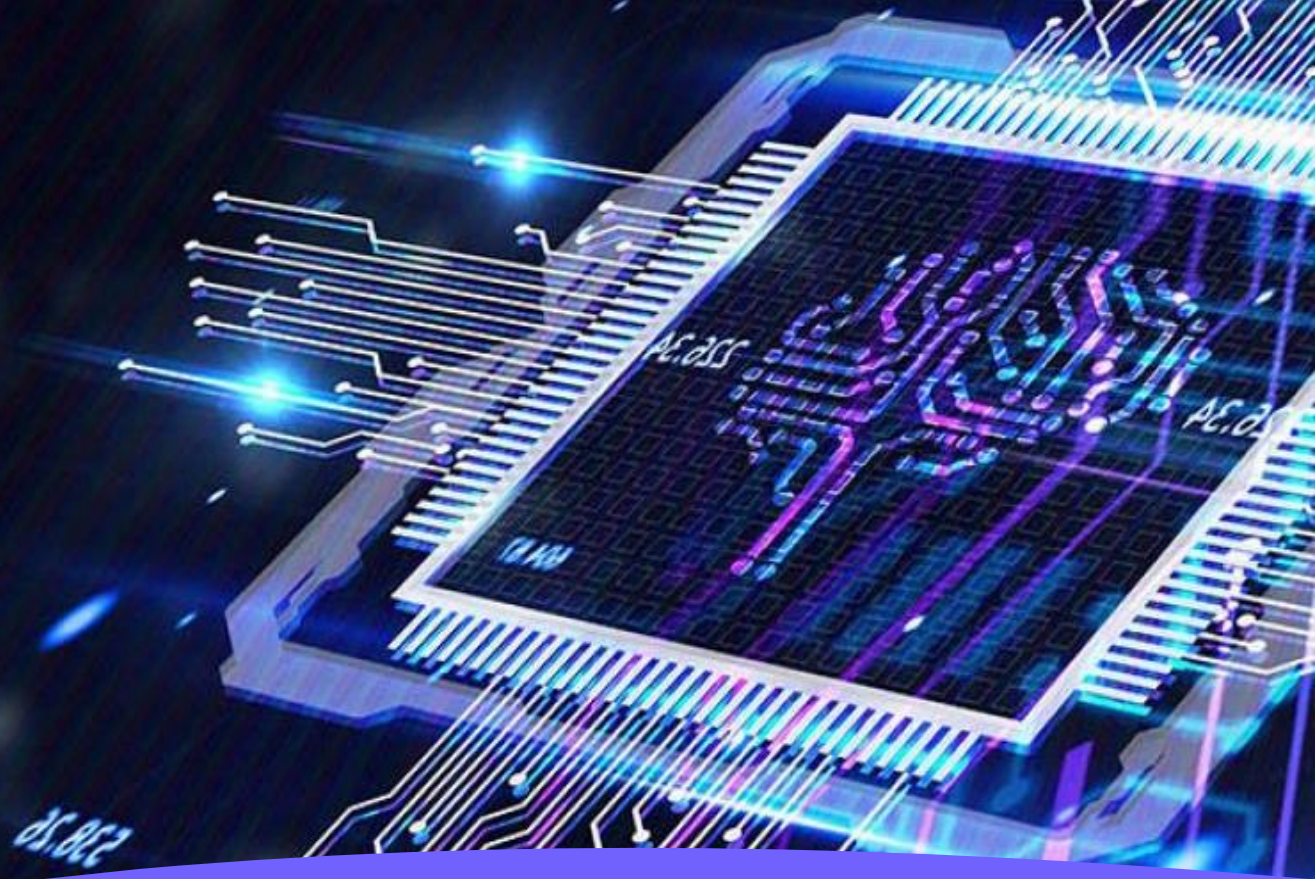




**FACULTAD
DE INGENIERIA**
Universidad de Buenos Aires



Aprendizaje de máquina II

Carrera de
Especialización en
Inteligencia Artificial

Agenda



- Versionado de modelos de ML utilizando MLflow

Versionado de modelos de ML



Motivación del versionado de modelos

El ciclo de vida de un proyecto de ciencia de datos puede ser prolongado y a lo largo de él, se pueden desarrollar **múltiples modelos** diferentes. Pueden cambiar las **arquitecturas**, los **hiperparámetros**, los **datos** de entrenamiento, los **desarrolladores**, la **infraestructura** en donde se los ejecuta, entre otras variables.

Es por eso que se necesita una manera de llevar un **registro ordenado** de los desarrollos realizados.



“
El **versionado de modelos de machine learning** se refiere al proceso de asignar versiones únicas y distintivas a los modelos de machine learning a medida que se desarrollan y mejoran a lo largo del tiempo. Cada vez que se realiza una actualización o modificación significativa en un modelo, se le asigna una nueva identificación para
”
distinguirlo de las versiones anteriores.

Seguimiento y documentación

Podemos generar un historial detallado de los cambios realizados, entender cómo ha sido la evolución a lo largo del tiempo e ir documentando los resultados obtenidos con cada modificación.

Reproducibilidad

Generar un control de versiones me permite reproducir los resultados obtenidos en el pasado. Se evita la pérdida de información y siempre se puede acceder a resultados previos.

Comparación y evaluación

Tener las versiones de un modelo centralizadas en un solo repositorio, me facilita la comparación entre las diferentes pruebas realizadas. Se pueden hacer pruebas con diferentes escenarios y facilitar la toma de decisiones basadas en datos.

Despliegue

El control de versiones me permite llevar un seguimiento de qué versiones se encuentran en cada etapa (producción, desarrollo, fuera de uso, etc.).

Ventajas del versionado de modelos

Versionado de modelos

MLflow es un producto de código abierto diseñado para administrar el ciclo de vida de desarrollo de aprendizaje automático.

Es una de las herramientas más utilizadas actualmente en la industria para versionar modelos de machine learning y posee integración directa con múltiples plataformas cloud.



neptune.ai



Weights & Biases

Versionado de modelos

MLflow es un producto de código abierto diseñado para administrar el ciclo de vida de desarrollo de aprendizaje automático.

Es una de las herramientas de machine learning y posee integraciones con:

Algunos links donde podemos encontrar referencias a herramientas utilizadas:

[Landscape](#)
[MLOps.toys](#)



versionar modelos de machine



neptune.ai



Weights & Biases

Versionado de modelos

MLflow es un producto de código abierto diseñado para administrar el ciclo de vida de desarrollo de aprendizaje automático.

Es una de las herramientas más utilizadas actualmente en la industria para versionar modelos de machine learning y posee integración directa con múltiples plataformas.

W&B es una herramienta que también permite almacenar el linaje de los artifacts.



Versionado de modelos

MLflow es un producto de código abierto diseñado para administrar el ciclo de vida de desarrollo de aprendizaje automático.

Es una de las herramientas más utilizadas actualmente en la industria para crear modelos de machine learning y posee integraciones con:

El uso on-premise de mlflow no permite utilizar el registro de modelos / ni el acceso multi-usuario. Aún así, vamos a usar mlflow para familiarizarnos ya que es una de las herramientas más utilizadas en la industria.

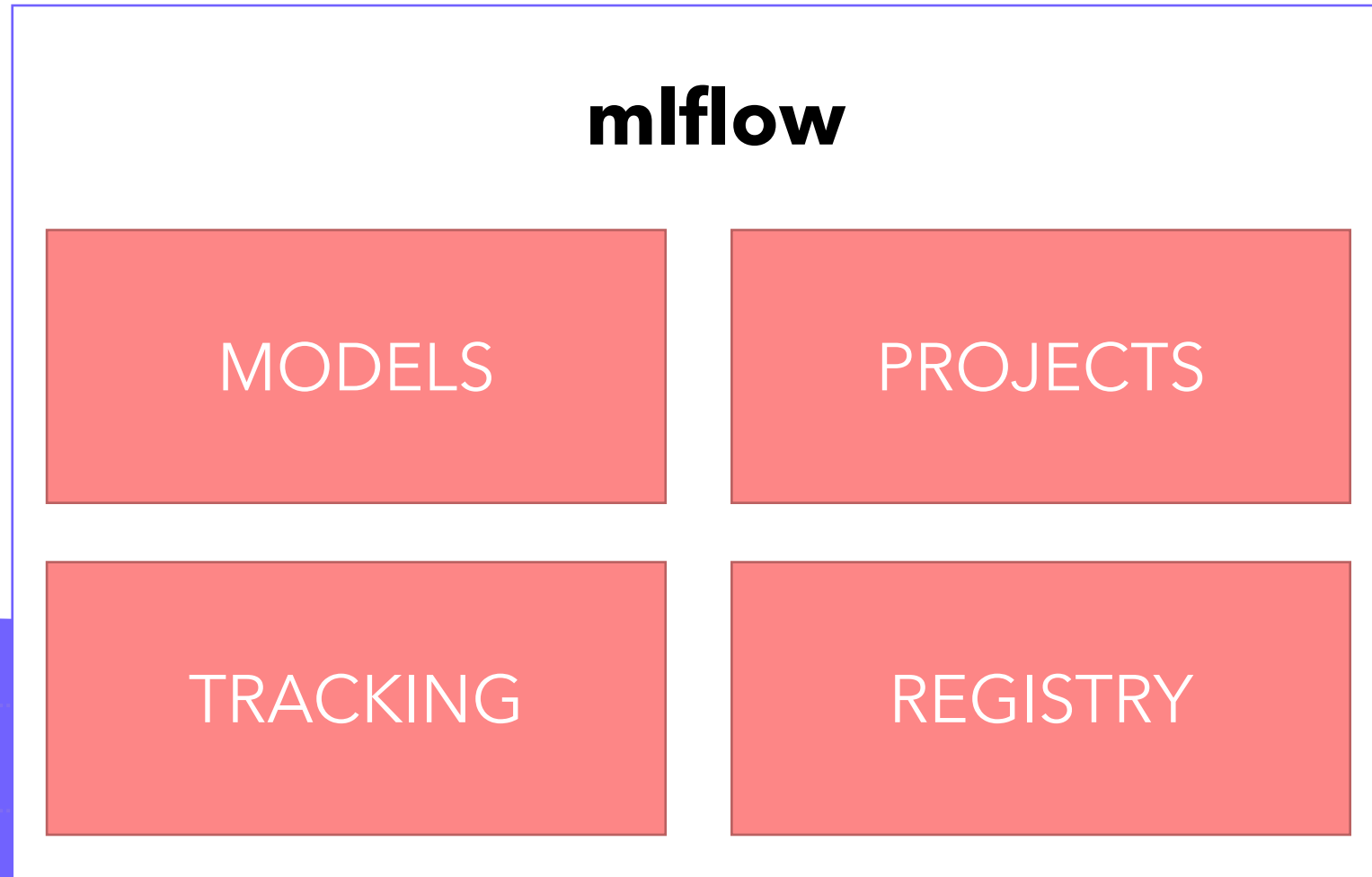


neptune.ai



Weights & Biases

¿Cómo está compuesto mlflow?



¿Cómo está compuesto mlflow?

Models

Permite empaquetar modelos en un mismo formato para facilitar la distribución. Esta funcionalidad permite que Mlflow funcione con múltiples bibliotecas como scikit-learn, Keras, MLlib, pyTorch, etc.

Projects

Nos brinda una manera de empaquetar código para lograr coherencia y reproducibilidad en los resultados obtenidos. Mlflow admite varios entornos para los proyectos como Conda, Docker, otros.

¿Cómo está compuesto mlflow?

Tracking

Este componente permite a los desarrolladores utilizar experimentos para registrar parámetros del modelo, versiones de código, métricas y las salidas de cada ejecución (run).

Un experimento es un conjunto de ejecuciones en las cuales entrenamos nuestros modelos de ML.

Registry

Este componente es un almacenamiento centralizado de modelos, define APIs y provee una UI para administrar de manera colaborativa el ciclo completo de un modelo de Mlflow. Nos permite acceder al linaje del modelo, la versión, estado y metadata.

Mlflow tracking



Sistema local de archivos

Útil para pruebas
personales



BBDD local + sist. de archivos

Útil para pruebas
personales y de
mediana escala



BBDD remota + datalake

Útil para sistemas en
producción