

# MLOps

Julio Mejia, Randi Paredes, Angel Crispin, Luis Cancino, Derian Herrera

1 de julio de 2021

## Resumen

*Los modelos de machine learning proporcionan insights valiosos al negocio, pero solo si esos modelos pueden acceder a los datos de la organización de forma continua y analizarlos. Las operaciones de machine learning (MLOps por sus siglas en inglés) son el proceso fundamental que lo hace posible.*

*Las MLOps son un proceso multifuncional, colaborativo e iterativo que pone en marcha las capacidades de data science. Para ello, las MLOps tratan el machine learning (ML) y otros tipos de modelos como artefactos de software reutilizables. Luego, los modelos se pueden implementar y monitorear continuamente mediante un proceso repetible.*

*Las MLOps ofrecen una integración continua y una implementación rápida y repetible de modelos. Como tal, ayudan a las empresas a descubrir información e insights valiosos a partir de sus datos con mayor rapidez. Las MLOps también incluyen el monitoreo y la repetición del entrenamiento continuo de modelos en producción para garantizar que funcionen de manera óptima a medida que los datos cambian (evolucionan) con el tiempo.*

## Abstract

*Machine learning models provide valuable information to the business, but only if those models can continuously access and analyze the organization's data. Machine learning operations (MLOps) are the fundamental process that makes this possible.*

*MLOps are a cross-functional, collaborative, and iterative process that puts data science capabilities to work. To do this, MLOps treat machine learning (ML) and other types of models as reusable software artifacts. The models can then be deployed and continuously monitored using a repeatable process.*

*MLOs offer continuous integration and fast, repeatable model deployment. As such, they help companies discover valuable information and insights from their data faster. MLOps also include continuous monitoring and re-training of models in production to ensure they perform optimally as data changes (evolves) over time.*

## I. INTRODUCCION

Los orígenes de MLOps se remontan a 2015 a partir de un artículo titulado "Deuda técnica oculta en sistemas de aprendizaje automático". Y desde entonces, el crecimiento ha sido particularmente fuerte. Tenga en cuenta que se espera que el mercado de soluciones MLOps alcance los \$4 mil millones para 2025..

## II. DESARROLLO

### II.1. ¿Qué es MLOps?

es inteligencia que proporciona un puente entre los científicos de datos y el equipo de producción. Conspira profundamente en la naturaleza y está diseñado para eliminar todos los residuos y hacer que el sistema de aprendizaje automático sea más escalable al proporcionar automatización y producir información altamente consistente del modelo de ML. MLOps es la idea de combinar la práctica establecida desde hace mucho tiempo de DevOps con el

campo emergente del aprendizaje automático. Es la creación de un entorno automatizado para el desarrollo de modelos, el reentrenamiento de modelos, la supervisión de derivas, la automatización de tuberías, el control de calidad y la gobernanza de un modelo en una sola plataforma. Dentro de MLOps existen tres niveles de implementación de Machine Learning:

- **Data:** datos, fase, ingestión, curado, etc.
- **Model:** testing, evaluación de los modelos, empaquetado y como se van a despegar.
- **Code:** el código, donde se ejecuta todo el modelo en sí.

MLOps = ML + DEV + OPS



## II.2. ¿Por qué MLOps?

La mayoría de los científicos de datos no son programadores de núcleo duro. Pueden crear el modelo más eficaz para un problema de aprendizaje automático, pero no tienen las habilidades para empaquetar, probar, implementar o mantener este modelo en producción. Se necesita alguien con el conocimiento de bases de datos, API de REST y una colección de otras habilidades de TI para hacerlo. Aquí es donde MLOps entra en escena. MLOps va más allá del desarrollo y diseño de modelos: también reúne la administración de datos, el desarrollo automatizado de modelos, el reentrenamiento de modelos, la generación de código, el desarrollo continuo y la supervisión del modelo. Al incorporar los principios de DevOps al aprendizaje automático, permite un ciclo de desarrollo más rápido, un mejor control de calidad y la capacidad de responder a los cambiantes requisitos empresariales.

## II.3. Beneficios de MLOps

Uno de los principales beneficios de las MLOps es que es compatible con tipos de modelos estadísticos, de data science y de aprendizaje automático, entre otros, para ofrecer un valor de negocios rápidamente. Para ello, las MLOps garantizan que los modelos se puedan implementar reiteradamente y monitorear de forma continua. El proceso de MLOps permite lo siguiente:

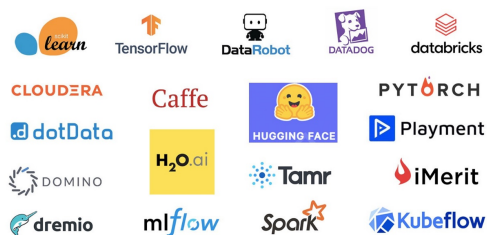
- Implementar más modelos con mayor rapidez con procesos automatizados
- Acelerar el tiempo de creación de valor con una rápida entrega de modelos
- Optimizar la productividad a través de la colaboración y la reutilización de modelos
- Reducir el riesgo de perder tiempo y dinero en modelos que nunca se incorporan a la producción
- Monitorear y actualizar continuamente modelos a medida que los datos evolucionan con el tiempo

## II.4. Desafíos de MLOps

Muchas organizaciones se enfrentan a desafíos al trasladar modelos de aprendizaje automático a entornos de producción. En promedio, entre el 60 % y el 80 % de los modelos creados con la intención de implementarlos nunca se llevan a cabo. Si implementas un modelo que se creó hace un periodo de seis meses a ocho meses, es posible que ese modelo ya esté obsoleto. Las organizaciones que tienen dificultades para integrar las aplicaciones de aprendizaje automático en las aplicaciones de producción existentes hacen perder tiempo y dinero en proyectos de data science que nunca se incorporan en la producción. Las MLOps pueden reducir en gran medida el riesgo de que se produzcan tales fallas y hacer que los modelos se incorporen a la producción con mayor rapidez, ya que, en definitiva, ahí es donde ofrezcan el mayor valor a un negocio.

## II.5. Herramientas

- **Kubeflow:** Inicialmente, Kubeflow fue iniciado por Google como una plataforma de código abierto para ejecutar Tensorflow usando Kubernetes. Pero ahora se ha expandido para ejecutar marcos multinube y multiarestela que requieren canalizaciones de ML.
- **Algorithmia:** Este famoso marco fue fundado por 2 desarrolladores en Washington DC en 2014. Esta plataforma cuenta actualmente con más de 60000 desarrolladores y más de 4500 algoritmos.
- **Pachyderm:** En esta plataforma, el proceso comienza con el control de versiones de datos combinado con la canalización de datos que da como resultado el linaje de datos y termina con la implementación de modelos de aprendizaje automático. También ofrece control de versiones de los datos para mantener los datos actualizados.
- **MLflow:** Automatiza el desarrollo del modelo y el modelo óptimo se puede seleccionar con mayor facilidad utilizando un servidor de seguimiento.
- **DataRobot:** La mejor parte de esta plataforma es su naturaleza ubicua. Puede acceder a DataRobot en cualquier lugar a través de cualquier dispositivo de múltiples maneras de acuerdo con las necesidades de su negocio.



## II.6. MLOps frente a DevOps frente a DataOps

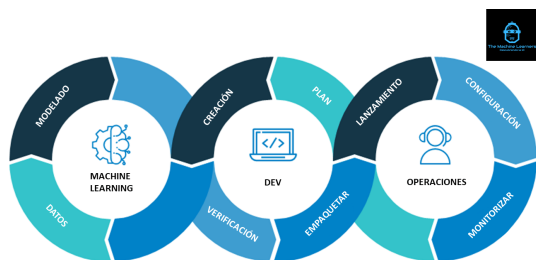
Las MLOps unifican la recopilación de datos, el procesamiento, el entrenamiento de modelos, la evaluación, la implementación y la repetición del entrenamiento en un solo proceso que los

equipos pueden mantener. Esta colaboración y comunicación entre DevOps, ITOps, ingenieros de datos, equipos de data science y otros departamentos aportan un entendimiento común de como se desarrollan y mantienen los modelos de aprendizaje automático en producción, de forma similar a lo que las DevOps hacen por el software. Las DevOps se centran en la entrega continua de software y en la automatización de la integración, las pruebas y la implementación de código. No involucran la gestión de datos o analítica. El proceso de MLOps se conforma a partir de las DevOps y se basa en la colaboración con los equipos de desarrollo para los servicios de implementación de modelos.

DataOps (operaciones de datos) se encarga de administrar pipelines de datos y automatizar procesos a fin de reducir el tiempo requerido para completar el análisis de datos.

## II.7. Principios de MLOps

- **Automatización:** Cuyas fases con proceso manual, ML pipeline y CI/CD pipeline.
- **Versiónado:** Basado en tres pilares, datos, modelo y código.
- **Testing:** base sobre la que va a funcionar todo el sistema.
- **Monitorización:** Los puntos clave a tener en cuenta son que hay que estar atentos a que no haya cambios en dependencias, hay que asegurar los datos invariantes, el sesgo en predicciones, un modelo "rancio" que no es capaz de responder a la suficiente velocidad, modelo numéricamente estable y degradación del modelo.
- **Reproductividad:** de la ingeniería de características, del entrenamiento del modelo y del despliegue del modelo. Hay que tener cuidado con la recogida de datos, para poder reproducir los resultados.
- **Herramientas:** mlflow, DVC, Kubeflow, ONNX, TensorFlow Extended, Apache Airflow, Cookiecutter, Metaflow, jupyter, PyTorch o Psycasfold.



### III. CONCLUSIONES

La configuración de MLOps puede ser bastante costosa, pero también ofrece importantes beneficios a las organizaciones que se toman en serio el aprendizaje automático reproducible y la mantenibilidad a largo plazo en un entorno de producción. Los flujos de trabajo de aprendizaje automático automatizados le permiten reciclar y volver a implementar su modelo. La integración continua mantiene el proceso ininterrumpido y mejora continuamente el modelo. El control de versiones de código y datos le permite recrear sus pruebas y revertir los resultados de producción a sus cambios originales. A medida que el aprendizaje automático crece como dominio, surgirán nuevos sistemas que facilitarán a su organización la configuración de MLOps. Con el tiempo, MLOps se volverá tan ubicuo e importante como DevOps.

### IV. RECOMENDACIONES

Existen varias plataformas MLOps para administrar el ciclo de vida del aprendizaje automático. Asegúrese de tener en cuenta los factores relevantes al seleccionar la plataforma.

### REFERENCIAS

- [1] neptune.ai.(2021) Best End-to-End MLOps Platforms - Leading Machine Learning Platforms That Every Data Scientist Need to Know. Recuperado de <https://n9.cl/a7fnz>
- [2] Alteryx. (2021). ¿Qué son las MLOps y cómo funcionan?. Recuperado de <https://n9.cl/dsg6r>
- [3] Medium. (2020). Introduction to MLOps - ILLUMINATION. Recuperado de <https://n9.cl/hh4or>
- [4] Towards Data Science. (2020). Get started with MLOps. Recuperado de <https://n9.cl/z3fnq>
- [5] Operaciones de Machine Learning (MLOps). (2021). Recuperado de <https://n9.cl/ha6qz>