

# MLOps

Jose Arias, Jose Gutierrez, Posi Vargas, Rodrigo Yanqui, Roby Zuñiga

30 de junio de 2021

## Resumen

*A medida que las empresas profundizan su compromiso con el aprendizaje automático, están descubriendo no solo sus beneficios, sino también sus desafíos. Las encuestas definitivas no parecen estar disponibles, pero numerosos líderes de opinión han notado que muchos modelos de ML (algunos dicen que llegan al 70-80 %) nunca pasan de la etapa de prototipo a la producción.*

*Una razón comúnmente citada para esta alta tasa de fallas es la dificultad de cerrar la brecha entre los científicos de datos que construyen y entrenan los modelos de inferencia y el equipo de TI que mantiene la infraestructura, así como los equipos de ingeniería de aprendizaje automático que desarrollan e implementan listos para producción. Aplicaciones de AA.*

*Examinaremos más de cerca los desafíos que impiden que el aprendizaje automático alcance su potencial de cambio total y discutiremos cómo una disciplina relativamente nueva, las operaciones de aprendizaje automático o MLOps, es la clave para optimizar el ciclo de vida de las aplicaciones de aprendizaje automático.*

## Abstract

*As companies deepen their commitment to machine learning, they are discovering not only its benefits, but also its challenges. Definitive surveys do not appear to be available, but many opinion leaders have noted that many ML models (some reaching 70-80%) never make it from prototype to production.*

*One reason cited for this high failure rate is the difficulty of bridging the gap between the data scientists who build and train the inference models and the IT team that maintains the infrastructure, as well as the machine learning engineering teams that develop and deploy ready for production. AA applications.*

*We'll take a closer look at the challenges preventing machine learning from reaching its full potential for change and discuss how one relatively discipline, Machine Learning Operations, or MLOps, is the key to optimizing the lifecycle of machine learning applications.*

## 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?

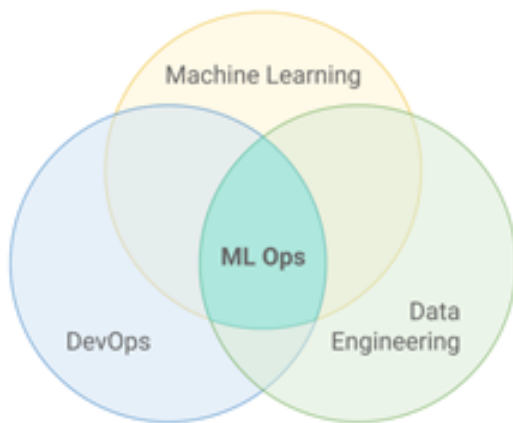
MLOps (Machine Learning Operations) es la práctica de combinar las lecciones aprendidas de DevOps para la producción de aprendizaje automático. Su función es llenar el vacío entre el científico de datos y los consumidores de aprendizaje automático.

MLOps es una extensión de la metodología DevOps que busca incluir activos de aprendi-

zaje automático y ciencia de datos como ciudadanos de primera clase dentro de la ecología DevOps. 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 desplegar.
- **Code:** el código, donde se ejecuta todo el modelo en sí.

MLOps = ML + DEV + OPS



## II.2. ¿Por qué empezar a aplicar MLOps?

En la actualidad, nos encontramos en un mundo orientado a datos, que está vinculado a la cantidad exponencialmente creciente de los mismos, recogidos digitalmente. Además, nos encontramos con la ascendente importancia de la Inteligencia Artificial y la Ciencia de Datos, que se deriva de esta tremenda cantidad de información generada.

Dependiendo de todos ellos, se pueden explorar de formas distintas, distinguiendo en capacidades (percepción, cognitivo y aprendizaje)

y casos de uso (visión, audio, voz y lenguaje natural).

## II.3. ¿Por qué MLOps?

La mayoría de los científicos de datos no son programadores empedernidos. 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 conocimiento de bases de datos, API REST y una colección de otras habilidades de TI para hacer eso. Aquí es donde entra en juego MLOps.

MLOps va más allá del desarrollo y diseño de modelos : también reúne la gestión de datos, el desarrollo automatizado de modelos, el re-entrenamiento de modelos, la generación de código, el desarrollo continuo y la supervisión del modelo . Al llevar 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 requisitos comerciales cambiantes.

## II.4. Beneficios de MLOps

- Modernización rápida mediante la colaboración con la gestión del ciclo de vida del aprendizaje automático
- Cree modelos y flujos de trabajo emulados
- Fácil uso de modelos de alta precisión
- Gestión eficaz de cada fase del ciclo de vida del aprendizaje automático
- También es útil para el sistema de gestión de recursos y su control.

## II.5. Desafíos de MLOps

- Despliegue y automatización
- Reproducibilidad de modelos y predicciones.
- Diagnósticos
- Gobernanza y cumplimiento normativo
- Escalabilidad
- Colaboración
- Usos comerciales

- Seguimiento y gestión

## II.6. Qué tener en cuenta para usar MLOps

- **Calidad de los datos:** tener en cuenta de dónde vienen, calidad, si son fiables, etc.
- **Degradación de los modelos:** al cabo del tiempo van perdiendo calidad.
- **Localidad:** en el momento de la preparación se están entrenando los modelos con unos datos específicos basados en una geografía.

## II.7. Cómo es el proceso de MLOps

- **Diseño:** ajuste de requerimientos, establecer las necesidades que tienen los usuarios y qué queremos cubrir, exploración de los datos, experimentación.
- **Desarrollo del modelo:** desarrollo de un modelo funcional capacitado para pasar a producción.
- **Operaciones:** despliegue, automatización de entrenamientos, extracción de datos, etc.

## II.8. ¿Cuáles son los Principios de MLOps?

- **Automatización:** cuyas fases con proceso manual, ML pipeline y CI/CD pipeline.
- **Versionado:** 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 Psycraft.



## III. CONCLUSIONES

La inversión de tiempo en la configuración de MLOps puede ser considerable, pero los beneficios también son considerables, para una organización que se toma en serio el aprendizaje automático repetible en producción y la mantenibilidad a largo plazo. Las canalizaciones de ML automatizadas le permiten volver a capacitar y volver a implementar modelos, y la integración continua le permite garantizar que los procesos no se rompan y que los modelos continúen mejorando. El control de versiones de los datos y el código garantiza que pueda reproducir experimentos y rastrear los resultados de producción hasta los cambios originales.

A medida que el aprendizaje automático madura como campo, aparecen nuevos sistemas que facilitan la configuración de MLOps para su organización. En poco tiempo, MLOps probablemente se considerará tan común e importante como DevOps.

## IV. RECOMENDACIONES

MLOps, o DevOps para Machine Learning, permite a los equipos de ciencia de datos y TI colaborar y aumentar el ritmo del desarrollo y la implementación de modelos a través de

la supervisión, validación y gobernanza de los modelos de Machine Learning.

### REFERENCIAS

- [1] Google Cloud. (s.f.). MLOps: canalizaciones de automatización y entrega continua en el aprendizaje automático. Recuperado de <https://cutt.ly/lmsHAYQ>
- [2] KeepCoding. (2019). ¿Qué es MLOps? [Guía completa]. Recuperado de <https://cutt.ly/lnod5w8>
- [3] Microsoft Azure. (2020). Operaciones de Machine Learning (MLOps). Recuperado de <https://cutt.ly/wmsHH5j>
- [4] Alteryx. (s.f.). Operaciones de Machine Learning (MLOps). Recuperado de <https://cutt.ly/cmsHVSc>
- [5] Towards Data Science. (2020). Get started with MLOps. Recuperado de <https://cutt.ly/HmsH8XN>