Acerca de Aprendizaje Automático en Producción - VISIT2024

Dr. Ricardo Rodriguez (Profesor Asociado) con la colaboración de la Lic. Leticia Rodriguez (Profesora Visitante)

Programa:

MLOps o Machine Learning Operations es una nueva área en el desarrollo de software que combina la Ingeniería del Software, la Ciencia de Datos y la Inteligencia Artificial.
Consiste en todas aquellas prácticas necesarias para planear, diseñar, modelar, implementar y monitorear sistemas de Inteligencia Artificial en producción.
Se entiende por *producción* a hacer disponible un software a sus consumidores y es parte del flujo del desarrollo de software que empresas públicas y privadas tienen como meta final en sus equipos.

Temario:

En los últimos años empezaron a surgir los primeros cursos y libros de la temática. Cuando un experto en Ciencia de Datos o Machine Learning se encuentra con su modelo terminado, listo para ser puesto como parte de un software productivo, la Ingeniería del Software juega un rol importante dónde tiempos de respuesta, volumen de datos, optimización de código, monitoreo continuo se vuelven fundamentales para el éxito del proyecto.
Muchas veces se requiere trabajar en el código del modelo para optimizarlo y mejorarlo para que el rendimiento que obtuvo en el mundo acotado del testeo, tenga valor en el universo más amplio del software productivo. Existen a su vez, necesidades propias del trabajo con datos y modelos que transforman la implementación productiva en un proceso iterativo de monitoreo y mejora continua de resultados.
A veces, no comprender la necesidad de la interacción de ambas disciplinas lleva a que muchas veces los modelos sean descartados o los costos de implementación sean demasiados altos. En el pasaje del mundo de la Ciencias de Datos a la Ingeniería del Software, la experiencia juega un rol central. Así se llega a esta nueva área que reúne la práctica y las particularidades del problema llamada MLOps.
Este curso busca que los estudiantes se acerquen a lo que ocurre con los modelos de Machine Learning una vez que están listos para ser productivos y como los procedimientos centrales del MLOps, contribuyen a que estos modelos estén disponibles de manera óptima, escalable y confiable.

Clase 1

- El ciclo de vida del desarrollo de software. Modelo Agile y Waterfall
- Desafíos y necesidades en el uso de la Inteligencia Artificial en el desarrollo de Software
- Desarrollo de modelos de Inteligencia Artificial: desde la recolección de datos hasta el modelo
- El ciclo de vida del desarrollo de software que incluye Aprendizaje Automático
- ¿Qué es DevOps? ¿Qué es MLOps?
- ¿Por qué MLOps es importante?
- Revisión de literatura científica relacionada con estos temas.

Clase 2

- Modelos Baseline y métricas. Entrenamiento de Modelos. Uso de paralelismo para el entrenamiento de modelos
- Monitoreo. Detección de drifts. Concepto de Monitoreo Continuo
- Concepto de Entrenamiento Continuo. Integración del Entrenamiento Continuo dentro del flujo automatizado del software.
- Revisión de literatura científica relacionada con estos temas.

Clase 3
- Construcción de Flujos Automatizados de Datos
- Construcción de Flujos Automatizados de Aprendizaje Automático
- Organización de entornos: desarrollo, testeo y producción
- Definición de CI/CD. Integración con Herramientas Git.
- Revisión de literatura científica relacionada con estos temas.

Clase 4
- Posibles arquitecturas de Software que usan modelos de IA.
- Aprovisionamiento de equipos. Escalabilidad y confiabilidad. Desarrollo en la nube: ventajas y desventajas.
- Herramientas específicas open source para Aprendizaje Automático en Producción: Kubernetes, Kubeflow, Tensorflow, Docker
- Revisión de literatura científica relacionada con estos temas.

Clase 5-7
- Trabajo práctico con Kubeflow y Tensorflow de Aprendizaje Automático en Producción
- Revisión de literatura científica relacionada con estos temas.

Clase 8
- MLOps para IA Generativa.
- Revisión de literatura científica sobre temas relacionados con Aprendizaje Automático en Producción fuera del alcance del curso.
- Cierre del curso

Bibliografía:

- D. Sculley et. al., **Hidden Technical Debt in Machine Learning Systems,** NeuIPS 2015
- D Kreuzberger et. al., **Machine Learning Operations (MLOps): Overview, Definition, and Architecture**. 2022
- Shreya Shankar, **Operationalizing Machine Learning: An Interview Study**. 2022
- Rob Ashmore, **Assuring the machine learning lifecycle: desiderata, methods, and challenges.** 2019
- Akshay Naresh Modi. **TFX: A TensorFlow-Based Production-Scale Machine Learning Platform**, KDD 2017

- Amreth Chandrasehar, **LLMOps: Evaluating and Fine Tuning LLM Models for Generative AI, International Journal of Machine Learning and Cybernetics (IJMLC), 1(1),** 2023, pp. 25-34
- Chip Huyen, **Designing Machine Learning Systems: An Iterative Process for Production-ready Applications**, O'Reilly Media, 2022
- Kubeflow https://www.kubeflow.org/
- Tensorflow Extended https://www.tensorflow.org/tfx?hl=es