

Guía docente

270968 - MLOPS - Sistemas de Aprendizaje Automático en la Producción (Mlops)

Última modificación: 23/11/2023

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 747 - ESSI - Departamento de Ingeniería de Servicios y Sistemas de Información.
Titulación: MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (Plan 2021). (Asignatura optativa).
Curso: 2023 **Créditos ECTS:** 6.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: SILVERIO JUAN MARTÍNEZ FERNÁNDEZ
Otros: Primer quadrimestre:
SANTIAGO DEL REY JUAREZ - 10
SILVERIO JUAN MARTÍNEZ FERNÁNDEZ - 10
MATÍAS SEBASTIÁN MARTÍNEZ MARTÍNEZ - 10

CAPACIDADES PREVIAS

Las dadas por las asignaturas de los cuatrimestres anteriores del máster. Fundamentos del aprendizaje automático.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa
CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos
CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

Genéricas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos
CG4. Diseñar y poner en marcha proyectos de ciencia de datos en dominios específicos de forma innovadora

Transversales:

CT2. Sostenibilidad y Compromiso Social. Conocer y comprender la complejidad de los fenómenos económicos y sociales típicos de la sociedad del bienestar; tener capacidad para relacionar el bienestar con la globalización y la sostenibilidad; lograr habilidades para utilizar de forma equilibrada y compatible la técnica, la tecnología, la economía y la sostenibilidad.
CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Básicas:

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

METODOLOGÍAS DOCENTES

Los contenidos teóricos de la asignatura se imparten en las clases de teoría. Estas clases se complementan con ejemplos prácticos y problemas que los estudiantes deben resolver en las horas de Aprendizaje Autónomo.

En las sesiones de laboratorio se consolidan los conocimientos adquiridos en las clases de teoría mediante la resolución de problemas y desarrollo de prácticas relacionados con los contenidos teóricos. Durante las clases de laboratorio, el profesor irá introduciendo nuevas técnicas y dejará una parte importante de la clase para que los estudiantes trabajen en los ejercicios propuestos.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1. Interpretar los conceptos básicos de la Ingeniería del Software para sistemas ML (machine learning), especialmente en relación al uso y explotación de prácticas MLOps.
2. Aplicar y analizar las prácticas de MLOps para construir modelos de ML, fomentando la reproducibilidad y la garantía de calidad.
3. Aplicar y analizar las prácticas de MLOps para implementar modelos de ML, fomentando el desarrollo de APIs y la entrega de componentes.
4. Describir los conceptos y métodos relacionados con el seguimiento de los datos obtenidos durante el uso de los sistemas de ML, con el fin de permitir ciclos de retroalimentación en respuesta a los cambios.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	100,0	64.94
Horas grupo grande	27,0	17.53
Horas grupo pequeño	27,0	17.53

Dedicación total: 154 h

CONTENIDOS

Conceptos básicos de Ingeniería de Software para sistemas ML (MLOps)

Descripción:

Motivación de la necesidad de ingeniería de software para sistemas ML. Introducción a MLOps y conceptos clave. Ingeniería de requisitos para ML. Plataformas de desarrollo colaborativo.

Prácticas de MLOps para construir modelos de ML

Descripción:

La complejidad y diversidad de los proyectos de ciencia de datos y los sistemas de ML exigen técnicas de ingeniería para garantizar que se construyan de manera sólida y preparada para el futuro. En este capítulo, abordamos las mejores prácticas de ingeniería de software para software de proyectos de ciencia de datos, incluidos los componentes de ML: sistemas de control de versiones; seguimiento y reproducibilidad de la canalización de ML; medición de software para ML; garantía de calidad para ML.

Prácticas de MLOps para implementar modelos de ML

Descripción:

La complejidad y diversidad de los sistemas de ML requieren técnicas de ingeniería para garantizar que se implementen de manera sólida y lista para la producción. En este capítulo, abordamos las mejores prácticas de ingeniería de software para componentes de ML: arquitectura de software para ML; implementar modelos de ML; API para ML; embalaje de componentes ML; automatización de canalizaciones de ML.

Monitoreo de datos obtenidos durante el uso de sistemas ML

Descripción:

Un problema clave en el desarrollo de software es la evolución del sistema ML en respuesta a las nuevas necesidades. El análisis de los datos obtenidos durante el uso del sistema ML por parte de sus usuarios, incluidos sus comentarios explícitos, permite conocer sus necesidades reales, de las que en ocasiones ni siquiera ellos son plenamente conscientes. Cada vez más nos encontramos con sistemas de software que necesitan ser conscientes de su contexto para prestar un correcto servicio. Esta restricción les obliga a monitorear los datos de contexto continuamente, descubrir cambios significativos y reaccionar en tiempo de ejecución (eventualmente, casi en tiempo real). Este tema describe el problema y revisa algunas técnicas básicas: monitoreo y telemetría; Ciclos MLOps y bucles de retroalimentación.

ACTIVIDADES

Estudio de conceptos básicos de Ingeniería de Software para sistemas ML (MLOps)

Objetivos específicos:

1

Competencias relacionadas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos
CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos
CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 5h 24m

Grupo grande/Teoría: 3h 36m

Aprendizaje autónomo: 1h 48m

Estudio de prácticas de MLOps para construir modelos ML

Objetivos específicos:

2

Competencias relacionadas:

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto
CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa
CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de dirección con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT2. Sostenibilidad y Compromiso Social. Conocer y comprender la complejidad de los fenómenos económicos y sociales típicos de la sociedad del bienestar; tener capacidad para relacionar el bienestar con la globalización y la sostenibilidad; lograr habilidades para utilizar de forma equilibrada y compatible la técnica, la tecnología, la economía y la sostenibilidad.
CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Dedicación: 10h 48m

Grupo grande/Teoría: 7h 12m

Aprendizaje autónomo: 3h 36m



Estudio de prácticas de MLOps para desplegar modelos ML

Objetivos específicos:

3

Competencias relacionadas:

CG4. Diseñar y poner en marcha proyectos de ciencia de datos en dominios específicos de forma innovadora

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Dedicación: 10h 48m

Grupo grande/Teoría: 7h 12m

Aprendizaje autónomo: 3h 36m

Estudio de conceptos para el seguimiento de datos obtenidos durante el uso de sistemas ML

Objetivos específicos:

4

Competencias relacionadas:

CG4. Diseñar y poner en marcha proyectos de ciencia de datos en dominios específicos de forma innovadora

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Dedicación: 10h 48m

Grupo grande/Teoría: 7h 12m

Aprendizaje autónomo: 3h 36m



Desarrollo práctico de un proyecto end-to-end de prácticas MLOps en el contexto de sistemas basados en ML

Descripción:

El alumno desarrollará progresivamente una práctica que le permita ejercitar los conceptos básicos introducidos en la parte teórica. Se desarrollará en equipos de 4-5 alumnos. El software resultante, debidamente documentado, se cargará en un repositorio de códigos. El equipo presentará un informe, escrito en inglés, resumiendo los principales aspectos de la práctica. Esto es, el proceso de construir e implementar un componente ML de un sistema basado en ML y una evaluación de la precisión de los modelos y algoritmos utilizados.

Objetivos específicos:

1, 2, 3, 4

Competencias relacionadas:

CG4. Diseñar y poner en marcha proyectos de ciencia de datos en dominios específicos de forma innovadora

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos

CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos

CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto

CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa

CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.

CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CT2. Sostenibilidad y Compromiso Social. Conocer y comprender la complejidad de los fenómenos económicos y sociales típicos de la sociedad del bienestar; tener capacidad para relacionar el bienestar con la globalización y la sostenibilidad; lograr habilidades para utilizar de forma equilibrada y compatible la técnica, la tecnología, la economía y la sostenibilidad.

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Dedicación: 95h 48m

Grupo pequeño/Laboratorio: 25h 12m

Aprendizaje autónomo: 70h 36m



Presentación del resumen de un artículo existente sobre MLOps

Descripción:

El alumno presentará el resumen de un artículo científico. Todos los estudiantes deben presentar (al menos) una vez. Los presentadores deben hacer al menos una pregunta a las otras presentaciones para fomentar las discusiones. Los profesores preparan una lista de artículos.

Objetivos específicos:

1

Competencias relacionadas:

CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos
CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos
CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

Dedicación: 8h 12m

Actividades dirigidas: 1h 48m

Aprendizaje autónomo: 6h 24m

Presentación del desarrollo práctico de un proyecto end-to-end de prácticas MLOps en el contexto de sistemas basados en ML

Objetivos específicos:

1, 2, 3, 4

Competencias relacionadas:

CG4. Diseñar y poner en marcha proyectos de ciencia de datos en dominios específicos de forma innovadora
CG3. Definir, diseñar e implementar sistemas complejos que cubran todas las fases en proyectos de ciencia de datos
CE5. Modelar, diseñar e implementar sistemas complejos de datos, incluyendo la visualización de datos
CE7. Identificar las limitaciones impuestas por la calidad de datos en un problema de ciencia de datos y aplicar técnicas para disminuir su impacto
CE10. Identificar los métodos de aprendizaje automático y modelización estadística a utilizar para resolver un problema específico de ciencia de datos y aplicarlos de forma rigurosa
CT3. TRABAJO EN EQUIPO: Ser capaz de trabajar como miembro de un equipo interdisciplinar ya sea como un miembro mas, o realizando tareas de direccion con la finalidad de contribuir a desarrollar proyectos con pragmatismo y sentido de la responsabilidad, asumiendo compromisos teniendo en cuenta los recursos disponibles.
CT5. TERCERA LENGUA: Conocer una tercera lengua, preferentemente el inglés, con un nivel adecuado oral y escrito y en consonancia con las necesidades que tendrán los titulados y tituladas.

CT2. Sostenibilidad y Compromiso Social. Conocer y comprender la complejidad de los fenómenos económicos y sociales típicos de la sociedad del bienestar; tener capacidad para relacionar el bienestar con la globalización y la sostenibilidad; lograr habilidades para utilizar de forma equilibrada y compatible la técnica, la tecnología, la economía y la sostenibilidad.

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CB8. Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

Dedicación: 8h 12m

Actividades dirigidas: 1h 48m

Aprendizaje autónomo: 6h 24m



SISTEMA DE CALIFICACIÓN

La nota se calcula ponderando la nota del proyecto (peso 90%) y la nota de una presentación teórica (peso 10%). Ambas actividades son obligatorias.

BIBLIOGRAFÍA

Básica:

- Hulten, Geoff. Building intelligent systems: a guide to machine learning engineering [en línea]. California: Apress, 2018 [Consulta: 15/01/2024]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pg-origsite=primo&docID=5357977>. ISBN 9781484234327.

Complementaria:

- Lanubile, Filippo; Martínez-Fernández, Silverio; Quaranta, Luigi. "Teaching MLOps in Higher Education through Project-Based Learning". 2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET) [en línea]. [Consulta: 15/01/2024]. Disponible a: <https://arxiv.org/pdf/2302.01048.pdf>.- Kästner, Christian. Machine Learning in Production [en línea]. Carnegie Mellon University, 2022 [Consulta: 15/01/2024]. Disponible a: <https://mlip-cmu.github.io/s2023/>.