

Guia docent

270968 - MLOPS - Sistemes d'Aprenentatge Automàtic a la Producció (Mlops)

Última modificació: 23/11/2023

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 747 - ESSI - Departament d'Enginyeria de Serveis i Sistemes d'Informació.
Titulació: MÀSTER UNIVERSITARI EN CIÈNCIA DE DADES (Pla 2021). (Assignatura optativa).
Curs: 2023 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

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

CAPACITATS PRÈVIES

Les donades per les assignatures dels quadrimestres anteriors del màster. Fonaments de l'aprenentatge automàtic.

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa
CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades
CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte

Genèriques:

CG3. Definir, dissenyar i implementar sistemes complexos que cobreixin totes les fases en projectes de ciència de dades
CG4. Dissenyar i posar en marxa projectes de ciència de dades en dominis específics de forma innovadora

Transversals:

CT2. Sostenibilitat i Compromís Social. Conèixer i comprendre la complexitat dels fenòmens econòmics i socials típics de la societat del benestar; tenir capacitat per relacionar el benestar amb la globalització i la sostenibilitat; obtenir habilitats per utilitzar de forma equilibrada i compatible la tècnica, la tecnologia, l'economia i la sostenibilitat.
CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.
CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

Bàsiques:

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.
CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.



METODOLOGIES DOCENTS

Els continguts teòrics de l'assignatura s'imparteixen a les classes de teoria. Aquestes classes es complementen amb exemples pràctics i problemes que els estudiants han de resoldre en les hores d'Aprenentatge Autònom.

En les sessions de laboratori es consoliden els coneixements adquirits a les classes de teoria mitjançant la resolució de problemes i desenvolupament de pràctiques relacionats amb els continguts teòrics. Durant les classes de laboratori, el professor anirà introduint noves tècniques i deixarà una part important de la classe per tal que els estudiants treballin en els exercicis proposats.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

1. Interpretar els conceptes bàsics de l'Enginyeria de Software per a sistemes ML, especialment en relació amb l'ús i aprofitament de les pràctiques MLOps.
2. Aplicar i analitzar les pràctiques MLOps per a construir models ML, fomentant la reproductibilitat i la garantia de qualitat.
3. Aplicar i analitzar pràctiques de MLOps per a desplegar models ML, fomentant el desenvolupament de API i el lliurament de components.
4. Descriure conceptes i mètodes relacionats amb la monitorització de les dades obtingudes durant l'ús dels sistemes de ML, amb la finalitat de permetre bucles de retroalimentació en resposta als canvis.

HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	100,0	64.94
Hores grup petit	27,0	17.53
Hores grup gran	27,0	17.53

Dedicació total: 154 h

CONTINGUTS

Conceptes bàsics d'Enginyeria del Software per a sistemes de ML (MLOps)

Descripció:

Motivació de la necessitat d'enginyeria del software per a sistemes ML. Introducció a MLOps i conceptes clau. Enginyeria de requisits per a ML. Plataformes de desenvolupament col·laboratiu.

Pràctiques MLOps per a construir models de ML

Descripció:

La complexitat i diversitat dels projectes de ciència de dades i els sistemes de ML requereixen tècniques d'enginyeria que garanteixin que es construeixen de manera robusta i preparada per al futur. En aquest capítol abordem les millors pràctiques d'enginyeria del software per al software de projectes de ciència de dades, inclosos els components de ML: sistemes de control de versions; reproductibilitat i seguiment de canalitzacions de ML; mesurament de software per a ML; garantia de qualitat per a ML.

Pràctiques de MLOps per a desplegar models de ML

Descripció:

La complexitat i diversitat dels sistemes de ML exigeixen tècniques d'enginyeria que garanteixin que es despleguen de manera robusta i llista per a la producció. En aquest capítol abordem les millors pràctiques d'enginyeria del software per a components de ML: arquitectura del software per a ML; desplegament de models de ML; API per a ML; empaquetat de components de ML; automatització de pipelines de ML.



Dades de seguiment obtinguts durant l'ús dels sistemes de ML

Descripció:

Un problema clau en el desenvolupament de software és l'evolució del sistema de ML en resposta a noves necessitats. L'anàlisi de les dades obtingudes durant l'ús del sistema de ML per part dels seus usuaris, inclosos els seus comentaris explícits, permet descobrir les seves necessitats reals, de les quals a vegades ni tan sols ells són plenament conscients. Cada vegada més ens trobem amb sistemes de software que necessiten conèixer el seu context per a prestar un servei correcte. Aquesta restricció els obliga a monitorar contínuament les dades de context, descobrir canvis significatius i reaccionar en temps d'execució (eventualment, gairebé en temps real). Aquest tema descriu el problema i repassa algunes tècniques bàsiques: monitoratge i telemetria; cicles MLOps i bucles de retroalimentació.

ACTIVITATS

Estudi dels conceptes bàsics de l'Enginyeria del Software per a sistemes de ML (MLOps)

Objectius específics:

1

Competències relacionades:

CG3. Definir, dissenyar i implementar sistemes complexos que cobreixin totes les fases en projectes de ciència de dades
CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades
CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

Dedicació: 5h 24m

Grup gran/Teoria: 3h 36m

Aprenentatge autònom: 1h 48m

Estudi de pràctiques MLOps per a construir models de ML

Objectius específics:

2

Competències relacionades:

CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte
CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa
CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.
CT2. Sostenibilitat i Compromís Social. Conèixer i comprendre la complexitat dels fenòmens econòmics i socials típics de la societat del benestar; tenir capacitat per relacionar el benestar amb la globalització i la sostenibilitat; obtenir habilitats per utilitzar de forma equilibrada i compatible la tècnica, la tecnologia, l'economia i la sostenibilitat.
CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.
CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

Dedicació: 10h 48m

Grup gran/Teoria: 7h 12m

Aprenentatge autònom: 3h 36m



Estudi de les pràctiques de MLOps per a desplegar models de ML

Objectius específics:

3

Competències relacionades:

CG4. Dissenyar i posar en marxa projectes de ciència de dades en dominis específics de forma innovadora

CG3. Definir, dissenyar i implementar sistemes complexos que cobreixin totes les fases en projectes de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

Dedicació: 10h 48m

Grup gran/Teoria: 7h 12m

Aprenentatge autònom: 3h 36m

Estudi de conceptes per a la supervisió de les dades obtingudes durant l'ús de sistemes de ML

Objectius específics:

4

Competències relacionades:

CG4. Dissenyar i posar en marxa projectes de ciència de dades en dominis específics de forma innovadora

CG3. Definir, dissenyar i implementar sistemes complexos que cobreixin totes les fases en projectes de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

Dedicació: 10h 48m

Grup gran/Teoria: 7h 12m

Aprenentatge autònom: 3h 36m



Desenvolupament pràctic d'un projecte end-to-end de pràctiques MLOps en el context de sistemes basats en ML

Descripció:

L'estudiant desenvoluparà progressivament una pràctica que li permetrà exercir els conceptes bàsics introduïts en la part teòrica. Es desenvoluparà en equips de 4-5 estudiants. El software resultant, degudament documentat, es carregarà a un repositori de codi. L'equip presentarà un informe, escrit en anglès, que resumirà els aspectes principals de la pràctica. Això és, el procés de construcció i desplegament d'un component ML d'un sistema basat en ML, i una avaluació de la precisió dels models i algorismes utilitzats.

Objectius específics:

1, 2, 3, 4

Competències relacionades:

CG4. Dissenyar i posar en marxa projectes de ciència de dades en dominis específics de forma innovadora

CG3. Definir, dissenyar i implementar sistemes complexos que cobreixin totes les fases en projectes de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o duent a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT2. Sostenibilitat i Compromís Social. Conèixer i comprendre la complexitat dels fenòmens econòmics i socials típics de la societat del benestar; tenir capacitat per relacionar el benestar amb la globalització i la sostenibilitat; obtenir habilitats per utilitzar de forma equilibrada i compatible la tècnica, la tecnologia, l'economia i la sostenibilitat.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autòdridigida o autònoma.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten- a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

Dedicació: 95h 48m

Grup petit/Laboratori: 25h 12m

Aprenentatge autònom: 70h 36m

Presentació d'un resum d'un article existent sobre MLOps

Descripció:

L'alumne presentarà el resum d'un article científic. Tots els alumnes han de presentar (almenys) una vegada. Els presentadors han de formular almenys una pregunta als altres per a fomentar el debat. Els professors preparen una llista d'articles.

Objectius específics:

1

Competències relacionades:

CG3. Definir, dissenyar i implementar sistemes complexos que cobreixin totes les fases en projectes de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

Dedicació: 8h 12m

Activitats dirigides: 1h 48m

Aprenentatge autònom: 6h 24m



Presentació del projecte de pràctiques MLOps en el context de sistemes basats en ML

Objectius específics:

1, 2, 3, 4

Competències relacionades:

CG4. Dissenyar i posar en marxa projectes de ciència de dades en dominis específics de forma innovadora

CG3. Definir, dissenyar i implementar sistemes complexos que cobreixin totes les fases en projectes de ciència de dades

CE5. Modelar, dissenyar i implementar sistemes complexos de dades, incloent-hi la visualització de dades

CE7. Identificar les limitacions imposades per la qualitat de dades en un problema de ciència de dades i aplicar tècniques per a disminuir el seu impacte

CE10. Identificar els mètodes d'aprenentatge automàtic i modelització estadística a utilitzar per resoldre un problema específic de ciència de dades, i aplicar-los de forma rigorosa

CT3. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip interdisciplinari, ja sigui com un membre més o dument a terme tasques de direcció, amb la finalitat de contribuir a desenvolupar projectes amb pragmatisme i sentit de la responsabilitat, tot assumint compromisos considerant els recursos disponibles.

CT5. TERCERA LLENGUA: Conèixer una tercera llengua, preferentment l'anglès, amb un nivell adequat oral i escrit i en consonància amb les necessitats que tindran els titulats i titulades.

CT2. Sostenibilitat i Compromís Social. Conèixer i comprendre la complexitat dels fenòmens econòmics i socials típics de la societat del benestar; tenir capacitat per relacionar el benestar amb la globalització i la sostenibilitat; obtenir habilitats per utilitzar de forma equilibrada i compatible la tècnica, la tecnologia, l'economia i la sostenibilitat.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB8. Que els estudiants sàpiguen comunicar les seves conclusions i els coneixements i raons darreres que les sustenten a públics especialitzats i no especialitzats d'una manera clara i sense ambigüitats.

Dedicació: 8h 12m

Activitats dirigides: 1h 48m

Aprenentatge autònom: 6h 24m

SISTEMA DE QUALIFICACIÓ

La nota es calcula ponderant la nota del projecte (pes 90%) i la nota de la presentació de l'article (pes 10%). Ambdues activitats són obligatòries.

BIBLIOGRAFIA

Bàsica:

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

Complementària:

- 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 línia]. [Consulta: 15/01/2024]. Disponible a: <https://arxiv.org/pdf/2302.01048.pdf>.- Kästner, Christian. Machine Learning in Production [en línia]. Carnegie Mellon University, 2022 [Consulta: 15/01/2024]. Disponible a: <https://mlip-cmu.github.io/s2023/>.