



Guia docent

300271 - BIGDATA - Macrodades i Minería de Dades

Última modificació: 09/06/2023

Unitat responsable: Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels
Unitat que imparteix: 701 - DAC - Departament d'Arquitectura de Computadors.

Titulació: MÀSTER UNIVERSITARI EN APLICACIONS I GESTIÓ DE L'ENGINYERIA DE TELECOMUNICACIÓ (MASTEAM) (Pla 2015). (Assignatura optativa).

Curs: 2023 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable:

Altres:

CAPACITATS PRÈVIES

Anglès, programació, probabilitat.

REQUISITS

Anglès, programació, probabilitat.

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Genèriques:

03 DIS. Dissenyar aplicacions d'alt valor afegit basades en les Tecnologies de la Informació i les Comunicacions (TIC), aplicades a qualsevol àmbit de la societat.

06 RES. Resoldre problemes i millorar processos en qualsevol àmbit social a partir de l'aplicació de les TIC, integrant coneixements de diversos àmbits i aplicant enginyeria d'alt nivell tecnològic.

Transversals:

05 TEQ. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip, ja sigui com un membre més, o realitzant 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.

03 TLG. TERCERA LLENGUA: Conèixer una tercera llengua, que serà preferentment l'anglès, amb un nivell adequat de forma oral i per escrit i amb consonància amb les necessitats que tindran les titulades i els titulats en cada ensenyament.

Bàsiques:

CB6. Tenir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context d'investigació.

METODOLOGIES DOCENTS

El curs s'organitza en activitats de programació en les quals l'estudiant desenvolupa projectes relacionats amb 'Big Data Analytics'. Usa metodologia d'aprenentatge basada en projectes.

OBJECTIUS D'APRENENTATGE DE L'ASSIGNATURA

En acabar el curs l'estudiant serà capaç d' aplicar una llista de tècniques de minería de dades per grans quantitats de dades, extreure informació útil a partir de grans volums de dades, programar usant la tècnica del 'map-reduce' i executar sobre granges d'ordenadors reals, virtuals o al núvol.



HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup petit	54,0	36.00
Hores aprenentatge autònom	96,0	64.00

Dedicació total: 150 h

CONTINGUTS

T1

Descripció:

Introducció al 'big data', presentació del curs, exemples d'aplicacions de la tecnologia de big data i de mineria de dades, presentació de l'entorn de desenvolupament i recursos de dades.

Activitats vinculades:

A1

Dedicació: 10h

Grup petit/Laboratori: 4h

Aprenentatge autònom: 6h

T2

Descripció:

Fonts de dades, sistemes de fitxers i bases de dades distribuïts, 'data streaming': tecnologies de indexació, memòria, fluxos de bits, bases de dades i evolució. Exemples de dades disponibles.

Activitats vinculades:

A1+A2

Dedicació: 25h

Grup petit/Laboratori: 15h

Aprenentatge autònom: 10h

T3

Descripció:

Processament i mineria de dades: fonaments bàsics i aplicacions del map-reduce, models d'aprenentatge (cerques, classificació, regressió, agrupació, extracció), inferència bayesiana, raonament lògic, incertesa i predicció.

Activitats vinculades:

A2

Dedicació: 115h

Grup petit/Laboratori: 35h

Aprenentatge autònom: 80h



ACTIVITATS

A1

Descripció:

Exercicis guiats: instal·lació de l'entorn de programació i de les eines de big data de Apache, exemples de programes bàsics (hello world, ús de llistes, diccionaris, etc.), preparació de les dades de prova i de les llibreries de aprenentatge artificial.

Material:

Atenea

Lliurament:

A1 in Atenea (30%)

Competències relacionades:

06 RES. Resoldre problemes i millorar processos en qualsevol àmbit social a partir de l'aplicació de les TIC, integrant coneixements de diversos àmbits i aplicant enginyeria d'alt nivell tecnològic.

Dedicació: 26h

Grup petit/Laboratori: 10h

Aprenentatge autònom: 16h

A2

Descripció:

Projecte: Classificació d'objectes basada en punts característics per mètodes variats. Ús d'arbres de decisió i xarxes Bayesianes per explicar causalitat. Predicció d'indicadors fent servir tècniques de regressió. Mostra i analitzar resultats a partir de reducció del dimensionament. Pre-processat, extracció i selecció de punts característics. Selecció del millors paràmetres per als models d'aprenentatge.

Material:

Atenea

Lliurament:

A2 in Atenea (70%)

Competències relacionades:

03 DIS. Dissenyar aplicacions d'alt valor afegit basades en les Tecnologies de la Informació i les Comunicacions (TIC), aplicades a qualsevol àmbit de la societat.

06 RES. Resoldre problemes i millorar processos en qualsevol àmbit social a partir de l'aplicació de les TIC, integrant coneixements de diversos àmbits i aplicant enginyeria d'alt nivell tecnològic.

05 TEQ. TREBALL EN EQUIP: Ser capaç de treballar com a membre d'un equip, ja sigui com un membre més, o realitzant 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.

03 TLG. TERCERA LLENGUA: Conèixer una tercera llengua, que serà preferentment l'anglès, amb un nivell adequat de forma oral i per escrit i amb consonància amb les necessitats que tindran les titulades i els titulats en cada ensenyament.

CB6. Tenir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context d'investigació.

Dedicació: 124h

Grup petit/Laboratori: 44h

Aprenentatge autònom: 80h

SISTEMA DE QUALIFICACIÓ

A1=30% + A2=70%



NORMES PER A LA REALITZACIÓ DE LES PROVES.

Cal portar ordinador personal a classe. Assistència obligatòria com a mínim al 80% de les classes. Les activitats es faran en grup.

BIBLIOGRAFIA

Bàsica:

- Géron, Aurélien. Hands-on machine learning with scikit-learn & tensorflow : concepts, tools, and techniques to build intelligent systems [en línia]. Sebastopol, CA: O'Reilly Media, Inc, [2017] [Consulta: 26/07/2022]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=4822582>. ISBN 9781491962299.

Complementària:

- Macías Lloret, Mario; Gómez Mauro; Tous Liesa, Rubén; Torres, Jordi. Introducción a Apache Spark : para empezar a programar el big data. Barcelona: UOC, 2015. ISBN 9788491160373.

- Mohanty, Hrushikesh; Bhuyan, Prachet; Chenthati, Deepak. Big Data : A Primer [en línia]. New Delhi: Springer India, 2015 [Consulta: 26/07/2022]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-81-322-2494-5>. ISBN 9788132224945.

- Leskovec, Jure; Rajaraman, Anand; Ullman, Jeffrey D. Mining of massive datasets [en línia]. 2nd ed. New York, N.Y. ; Cambridge: Cambridge University Press, 2014 [Consulta: 26/07/2022]. Disponible a: <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=807230>. ISBN 9781107077232.

- Garreta, Raúl; Moncecchi, Guillermo. Learning scikit-learn : machine learning in Python. Birmingham: Packt Publishing, 2013. ISBN 9781783281930.

RECURSOS

Altres recursos:

Atenea