

300271 - BIGDATA - Macrodades i Minería de Dades

Unitat responsable: 300 - EETAC - Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels
Unitat que imparteix: 701 - AC - Departament d'Arquitectura de Computadors
Curs: 2016
Titulació: MÀSTER UNIVERSITARI EN APLICACIONS I GESTIÓ DE L'ENGINYERIA DE TELECOMUNICACIÓ (MASTEAM) (Pla 2015). (Unitat docent Optativa)
Crèdits ECTS: 6 Idiomes docència: Anglès

Professorat

Responsable: Cristina Barrado
Altres: Cristina Barrado
Pablo Royo

Capacitats prèvies

Anglès, programació, probabilitat.

Requisits

x

Competències de la titulació a les quals contribueix l'assignatura

Bàsiques:

CB6. 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.

CB7. CB7 - Que els estudiants sàpiguen aplicar els coneixements adquirits i la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB10. CB10 - Que els estudiants tinguin les habilitats d'aprenentatge que els permetin seguir estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

Específiques:

11 MTM. Planificar i executar un projecte de desenvolupament d'una aplicació de les TIC a un procés nou, o de millora d'un ja existent, en qualsevol àmbit de la vida social.

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.

Transversals:

05 TEQ N3. TREBALL EN EQUIP - Nivell 3: Dirigir i dinamitzar grups de treball, resolent-ne possibles conflictes, valorant el treball fet amb les altres persones i avaluant l'efectivitat de l'equip així com la presentació dels resultats generats.

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.

06 URI N3. ÚS SOLVENT DELS RECURSOS D'INFORMACIÓ - Nivell 3: Planificar i utilitzar la informació necessària per a un treball acadèmic (per exemple, per al treball de fi de grau) a partir d'una reflexió crítica sobre els recursos d'informació utilitzats.

300271 - BIGDATA - Macrodades i Minería de Dades

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 totals de dedicació de l'estudiantat

Dedicació total: 75h	Hores grup petit:	27h	36.00%
	Hores aprenentatge autònom:	48h	64.00%

300271 - BIGDATA - Macrodades i Minería de Dades

Continguts

<p>T1</p>	<p>Dedicació: 10h Grup petit: 4h Aprentatge autònom: 6h</p>
<p>Descripció: Introducció al 'big data', presentació del curs, exemples d'aplicacions de la tecnologia de big data i de minería de dades, presentació de l'entorn de desenvolupament i recursos de dades.</p> <p>Activitats vinculades: A1</p> <p>Objectius específics: x</p>	
<p>T2</p>	<p>Dedicació: 25h Grup petit: 10h Aprentatge autònom: 15h</p>
<p>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.</p> <p>Activitats vinculades: A1+A2</p> <p>Objectius específics: x</p>	
<p>T3</p>	<p>Dedicació: 115h Grup petit: 35h Aprentatge autònom: 80h</p>
<p>Descripció: Processament i minería de dades: fonaments bàsics i aplicacions del map-reduce, models d'aprenentatge (cerques, classificació, regressió, agrupació, extracció), inferència bayesiana, raonament lògic, incertesa i predicció.</p> <p>Activitats vinculades: A2</p> <p>Objectius específics: x</p>	

300271 - BIGDATA - Macrodades i Minería de Dades

Planificació d'activitats

A1	Dedicació: 26h Grup petit: 10h Aprenentatge autònom: 16h
<p>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.</p> <p>Material de suport: Atenea</p> <p>Descripció del lliurament esperat i vincles amb l'avaluació: A1 Atenea (30%)</p> <p>Objectius específics: x</p>	
A2	Dedicació: 124h Grup petit: 44h Aprenentatge autònom: 80h
<p>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.</p> <p>Material de suport: Atenea</p> <p>Descripció del lliurament esperat i vincles amb l'avaluació: A2 Atenea (70%)</p> <p>Objectius específics: x</p>	

Sistema de qualificació

A1=30% + A2=70%

300271 - BIGDATA - Macrodades i Minería de Dades

Bibliografia

Bàsica:

Macias 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.

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

Complementària:

Mohanty, Hrushiksha; Bhuyan, Prachet; Chenthati, Deepak. Big Data [Recurs electrònic]: A Primer [en línia]. New Delhi: Springer, 2015. Disponible a: <<http://dx.doi.org/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. Disponible a: <<http://site.ebrary.com/lib/upcatalunya/docDetail.action?docID=10514268>>. ISBN 9781107077232.

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction. 2nd ed. New York [etc.]: Springer, 2009. ISBN 9780387848570.

Provost, Foster; Fawcett, Tom. Data science for business : what you need to know about data mining and data-analytic thinking. Sebastopol, CA: O'Reilly, 2013. ISBN 9781449361327.

Altres recursos:

Material informàtic

Atenea course resources

Recurs