



Universitat de Lleida

GUIA DOCENT
EXPLOTACIÓ DE DADES

Coordinació: BEJAR TORRES, RAMON

Any acadèmic 2022-23

Informació general de l'assignatura

Denominació	EXPLOTACIÓ DE DADES			
Codi	103089			
Semestre d'impartició	1R Q(SEMESTRE) AVALUACIÓ CONTINUADA			
Caràcter	Grau/Màster	Curs	Caràcter	Modalitat
	Màster Universitari en Enginyeria Informàtica	2	OPTATIVA	Presencial
Nombre de crèdits assignatura (ECTS)	6			
Tipus d'activitat, crèdits i grups	Tipus d'activitat	PRALAB	TEORIA	
	Nombre de crèdits	3	3	
	Nombre de grups	1	1	
Coordinació	BEJAR TORRES, RAMON			
Departament/s	INFORMÀTICA I ENGINYERIA INDUSTRIAL			
Distribució càrrega docent entre la classe presencial i el treball autònom de l'estudiant	30% del temps en sessions presencials (3 hores/setmana) i 70% del temps en treball autònom			
Informació important sobre tractament de dades	Consulteu aquest enllaç per a més informació.			
Idioma/es d'impartició	Anglès			

Professor/a (s/es)	Adreça electrònica professor/a (s/es)	Crèdits impartits pel professorat	Horari de tutoria/lloc
BEJAR TORRES, RAMON	ramon.bejar@udl.cat	3	
MATEU PIÑOL, CARLOS	carles.mateu@udl.cat	3	

Informació complementària de l'assignatura

Per poder cursar la assignatura, el estudiant ha de tenir **coneixements sòlids y experiència previa en programació estructurada, orientada a objectes i funcional amb el llenguatge python**

Objectius acadèmics de l'assignatura

1. Conèixer les eines actuals per a la neteja i anàlisi de dades
2. Conèixer les bases per al desenvolupament de procediments centrats en dades utilitzant eines de programació interactives
3. Conèixer com transformar les dades en brut de qualsevol font de dades consistent per a la seva anàlisi
4. Conèixer com implementar i depurar procediments per a la transformació de conjunts de dades massius utilitzant enfocaments Big Data
5. Adquirir un esperit escèptic davant de conjunts de dades per incentivar l'anàlisi exploratòria utilitzant eines informàtiques
6. Adquirir les eines i coneixements per a l'anàlisi descriptiva de sèries potencialment massives de dades i intractables
7. Conèixer i utilitzar la major quantitat d'algoritmes bàsics de mineria de dades per descobrir les característiques rellevants de grans conjunts de dades
8. Conèixer i utilitzar algorismes bàsics i avançats per aprenentatge automàtic adequat per a aplicacions de dades grans
9. Conèixer els principis bàsics del sistemes de recomanació

Competències

Competències estratègiques de la Universitat de Lleida

- **UdL2:** Domini d'una llengua estrangera.

Competències transversals de la EPS

- **EPS1:** Capacitat de planificació i organització del treball personal.
- **EPS3:** Capacitat per transmetre informació, idees, problemes i solucions alhora un especialitzat i sense públic especialitzat.
- **EPS4:** Capacitat per concebre, dissenyar i implementar projectes i / o contribuir a noves solucions, utilitzant eines d'enginyeria.
- **EPS5:** Estar motivat per la qualitat i millora constant.

Competències generals

- **CG4:** Capacitat per matemàticament modelitzar, calcular i simular en empreses tecnològiques i centres d'enginyeria, en particular pel que fa a les tasques de recerca, desenvolupament i innovació en tots els camps relacionats amb l'enginyeria informàtica.
- **CG8:** Capacitat per aplicar els coneixements adquirits per resoldre problemes en situacions noves i no familiars i en contextos multidisciplinars més amplis i més, i ser capaç d'integrar aquest coneixement.

Competències bàsiques

- **CB3:** Els estudiants són capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis basats en informació que era incompleta o limitada, inclogui reflexions sobre les social i ètica.
- **CB4:** Els estudiants poden comunicar les seves conclusions i els coneixements i raons últimes que les sustenten, a públics especialitzats i no especialitzats de manera clara i inequívoca.

Competències específiques del grau

- **CE1:** Capacitat per comprendre i aplicar els coneixements avançats en computació d'alt rendiment i mètodes numèrics i computacionals a problemes d'enginyeria.
- **CE4:** Capacitat per a modelar, dissenyar, definir l'arquitectura, implantar, gestionar, operar, administrar i mantenir aplicacions, xarxes, sistemes, serveis i continguts informàtics.

Continguts fonamentals de l'assignatura

1. Data mining i aprenentatge automàtic
 - Extracció eficient de elements distintius i freqüents. Extracció de subconjunts freqüents
 - Anàlisi i Reducció de dimensionalitat
 - Regressió lineal i logística and algorismes basats en SGD
 - Clasificadors Naive Bayes
 - Clustering
 - Sistemes de recomanació
2. Xarxes neuronals i Deep Learning
 - Introducció a les xarxes neuronals
 - Xarxes neuronals profundes
 - Xarxes Convolucionals
 - Xarxes Recurrents
3. Control: Reinforcement Learning

Eixos metodològics de l'assignatura

Cada setmana, cada estudiant rebrà:

- Tres hores d'assistència a classe. Les classes es duen a terme mitjançant explicacions teòriques acompanyades d'exemples il·lustratius a la primera part, finalitzant amb exercicis pràctics en la segona part. Com a material de suport de la classe, seguirem les diapositives o quaderns python del curs.
- Altres materials de suport per seguir el tema d'una manera no presencial.

L'avaluació és contínua al llarg del semestre i consta de quatre parts diferents:

- Dues pràctiques: l'extensió de l'aplicació bigdata van començar en el tema anterior, "processament massiu de dades» amb xarxes neuronals i eines de dades grans.
- Dos informes i presentacions orals sobre la integració de les xarxes neuronals i les dades mining eines en la seva aplicació bigdata.

Degut al COVID-19 algunes classes seran en format no presencial, és a dir, mitjançant vídeos i videoconferències.

Pla de desenvolupament de l'assignatura

Setmanes:

1. Mining frequent items/item sets.
2. Mining distinct elements.
3. Dimensionality reduction.
4. Linear and Logistic regression.
5. Naive Bayes classifiers.
6. Clustering - crisp and probabilistic.
7. Recommender systems.
8. Work on Data mining projects.
9. Work on Big data application.
10. Neural Networks.
11. Feed forward neural networks.
12. Deep Learning.
13. Convolutional Neural Networks.
14. Convolutional Neural Networks II.
15. Recursive Neural Networks.
16. Control: Reinforcement Learning.
17. Work on Deep Learning projects.
18. Work on Big Data Applications.
19. Final presentations of big data projects.

Degut al COVID-19 algunas classes seran en format no presencial, és a dir, mitjançant vídeos i videoconferències.

Sistema d'avaluació

L'avaluació d'aquest curs es basa en l'avaluació contínua.

Activitats d'avaluació	%	Dates	O/V (1)	I/G (2)
<i>Data mining application</i> (3)	40%	End Sem.	M	G
<i>Data mining application presentation</i> (4)	10%	Middle Sem.	M	G
Deep learning exercises	25%	Middle Sem.	M	G
Data mining exercises	25%	Middle Sem.	M	G

(1) Mandatory / Voluntary

(2) Individual / Group

(3) Hi haurà una revisió del treball realitzat per cada membre del grup

(4) Cada membre del grup ha de participar igualment en la presentació i respondre preguntes de l'avaluador

Bibliografia i recursos d'informació

- Wes McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython. O'Really, 2012
- Holden Karau, Andy Konwinski, Patrick Wendell & Matei Zaharia. Learning Spark. O'Really, 2015
- Jure Leskovec, Anand Rajaraman & Jeffrey David Ullman. Mining of Massive Datasets. Cambridge University Press , 2014 (El podeu trobar a <http://www.mmds.org/> o també a la biblioteca de la universitat).