

## GRAU EN ENGINYERIA INFORMÀTICA DE GESTIÓ I SISTEMES D'INFORMACIÓ

### 109810 - BIG DATA

#### Informació general

- Tipus d'assignatura : Optativa
- Coordinador : Alfons Palacios González
- Curs: Quart
- Trimestre: Segon
- Crèdits: 6
- Professorat:
  - Sandra Obiol Madrid <[sobiol@tecnocampus.cat](mailto:sobiol@tecnocampus.cat)>

#### Idiomes d'impartició

- Castellà

Curs impartit en Castellà

#### Competències que es treballen

##### Bàsica

- B2\_ Que els estudiants sàpiguen aplicar els seus coneixements a la seva feina o vocació d'una forma professional i tinguin les comptències que demostren mitjançant l'elaboració i defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi
- B3\_ Que els estudiants tinguin la capacitat de reunir i interpretar dades rellevants (normalment dins de la seva àrea d'estudi), per emetre judicis que incloguin una reflexió sobre temes rellevants de caire social, científica o ètica
- B4\_ Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tan especialitzat com no especialitzat
- B5\_ Que els estudiants hagin desenvolupat aquelles habilitats d'aprenentatge necessàries per emprendre estudis posteriors amb un alt grau d'autonomia

##### Comú

- CIN15\_ Coneixement i aplicació dels principis fonamentals i tècniques bàsiques dels sistemes intel·ligents i la seva aplicació pràctica

##### Específica

- EFB1\_ Capacitat per a la resolució dels problemes matemàtics que puguin plantejar-se a l'enginyeria. Aptitud per aplicar els coneixements sobre: àlgebra lineal, càlcul diferencial i integral, mètodes numèrics, algorítmica numèrica, estadística i optimització

##### Transversal

- T1\_ Que els estudiants coneixin un tercer idioma, que serà preferentment l'anglès, amb un nivell adequat de forma oral i per escrit, d'acord amb les

necessitats que tindran les graduades i els graduats a cada titulació

- T2\_ Que els estudiants tinguin capacitat per a treballar com a membres d'un equip interdisciplinari 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, assumint compromisos tenint en compte els recursos disponibles

## Descripció

---

L'objectiu del curs es dominar eines analítiques aplicades a entorns Big Data. Entendre el rol de la indústria 4.0 (I4.0) com un nou disruptor de l'espai competitiu. Els tres pilars principals de I4.0 són tecnologia de sensors + connectivitat, producció + robòtica i computació + big data. Aquest últim serà cobert en profunditat. Aquesta matèria és essencial pels enginyers per ajudar-los a crear valor a través del coneixement que un pot extreure dels entorns Big Data.

La major part de les sessions cobriran eines de Machine Learning aplicades a context I4.0 i casos reals. El procés d'examinar grans quantitats de dades i de diferent naturalesa per descobrir patrons ocults, obtenir noves perspectives i possiblement informació rellevant i útil, serà explicat detalladament. El curs descriurà en la part final alguns dels últims avenços en Deep Learning (GAN) i Reinforcement Learning molt aplicats en la indústria

## Resultats d'aprenentatge

---

A nivell general, aquesta assignatura contribueix als següents resultats d'aprenentatge especificats per a la matèria optativa a la qual pertany

A un nivell més concret, en acabar l'assignatura l'estudiant ha de ser capaç de:

Ra1: Entendre el rol de la Indústria 4.0 i el Big Data per revelar nou coneixement.

Ra2: Preparar dades per resoldre problemes complexos

Ra3: Visualitzar i generar informes

Ra4: Aplicar tècniques analítiques d'aprenentatge Supervisat

Ra5: Entendre els diferents Mètodes d'aprenentatge No Supervisat.

Ra6: Avaluar Models

## Metodologia de treball

---

Tots els conceptes teòrics de la matèria s'exposaran en classes de teoria (grups grans) i/o en sessions de laboratori (grups petits). En aquestes classes, i a discreció dels docents impartidors, també es resoldran exercicis i problemes de caire més pràctic. Així mateix i sempre a discreció dels impartidors, es podrà demanar als estudiants que resolguin, de manera individual o en grup, problemes i/o exercicis breus. Aquestes activitats, que per la seva naturalesa d'optativitat i brevetat serviran a l'estudiant com a instrument d'autoavaluació del seu assoliment dels continguts de la matèria i podran ser utilitzades per part del docent per valorar-la.

Els conceptes de caire més pràctic i tot el que en essència es pugui considerar l'aplicació pràctica dels conceptes teòrics seran treballats en grups petits (de laboratori). En les sessions que es programin a aquest efecte es donaran les eines escaients per a resoldre les activitats programades. En ocasions els estudiants les hauran de finalitzar durant el temps d'aprenentatge autònom. Sempre que es consideri escaient es posarà a disposició dels estudiants activitats de caire totalment opcional que l'ajudin a preparar i a preparar-se per a les de caire obligatori.

## Continguts

---

### PART I

- 1 Història de la ciència de dades. Del Business Intelligence al Big Data
- 2 Qualitat i visualització de les dades. Informes i quadres de comandament
- 3 Classificació

- 3.1 GLM
  - 3.2 Arbres
  - 3.3 Altres mètodes
- PART II
- 4 Mètodes Clustering
    - 4.1 Mesures de distància
    - 4.2 Kmeans
    - 4.3 Clustering jeràrquic
  - 5 Anàlisi de text
  - 6 Avaluació del model
  - 7 Projecte

## Activitats d'aprenentatge

---

Es posa a disposició dels estudiants tot un seguit d'activitats de caire eminentment pràctic (exercicis curts, problemes...) que són la base de les activitats d'aprenentatge de l'assignatura. Aquestes activitats els estudiants/es les hauran de resoldre, sovint de manera no presencial, seguint les indicacions dels docents i també seran treballades a classe, ja sia com a exemples en les sessions de teoria, ja sia en les sessions de laboratori. Si bé aquestes activitats tindran caràcter optatiu (els docents no en verificaran de manera individualitzada la realització per part dels estudiants), seran imprescindibles per assolir els coneixements teórico-pràctics de l'assignatura.

Amb l'objectiu de recollir evidències de l'assoliment dels resultats d'aprenentatge esperats es realitzen les següents activitats de caràcter avaluatiu:

PeX = Prova escrita (Examen) [Relacionada amb totes les competències]

La prova inclourà els continguts associats a tots els resultats d'aprenentatge

PLab = Pràctiques Laboratori [Relacionada amb totes les competències]

Les pràctiques permetran a l'estudiant practicar conceptes descrits a teoria

Evidència dels resultats d'aprenentatge: Ra2, Ra3, Ra4, Ra5 i Ra6

Proj = Presentació Projecte Big Data [Relacionada amb totes les competències]

Els estudiants presentaran un projecte descrivint totes i cadascuna de les etapes que han desenvolupat. Es lliurarà el codi, el document de treball i la presentació realitzada

Evidència dels resultats d'aprenentatge: Tots

ExiP = Exercicis i participació a classe (presentació de la implementació d'un dels algorismes explicats a classe) [Relacionada amb totes les competències]

Els estudiants i/o grups que participin en la resolució de problemes

Evidència dels resultats d'aprenentatge: Tots

Observacions:

Per superar les activitats avaluatives, els estudiants hauran de demostrar el Nivell MECES - 2:

- (punt c) tenir la capacitat de recopilar i interpretar dades i informacions sobre les que fonamentar les seves conclusions incloent-hi, quan calgui i sigui pertinent, la reflexió sobre assumptes d'índole social, científica o ètica en l'àmbit del seu camp d'estudi
- (punt e) saber comunicar a tot tipus d'audiències (especialitzades o no) de manera clara i precisa, coneixements, metodologies, idees, problemes i solucions en l'àmbit del seu camp d'estudi;
- (punt f) ser capaços d'identificar les seves pròpies necessitats formatives en el seu camp d'estudi i entorn laboral o professional i d'organitzar el seu propi aprenentatge amb un alt grau d'autonomia en tot tipus de contextos

Per a cada activitat, els docents n'informaran de les normes i condicions particulars que les regeixin

Les activitats unipersonals pressuposen el compromís de l'estudiant de realitzar-les de manera individual i sense cap mena de col·laboració amb d'altres persones. Es consideraran suspeses (qualificació 0) totes aquelles activitats en què l'estudiant no s'ajusti a aquest compromís d'individualitat, independentment del seu paper (emissor o receptor) i sense que això exclouï la possible aplicació d'altres sancions d'acord amb el Règim Disciplinari vigent.

Igualment, les activitats que s'hagin de realitzar en grup pressuposen el compromís per part dels estudiants que l'integren de realitzar-les en el si del grup i sense cap mena de col·laboració amb d'altres grups o persones que en siguin alienes (individualitat grupal). Es consideraran suspeses (qualificació 0) totes aquelles activitats en què el grup no hagi respectat aquest compromís amb independència del seu paper (emissor o receptor) i sense que això exclouï la possible aplicació d'altres sancions d'acord amb el Règim Disciplinari vigent.

En el cas d'activitats que puguin fer-se en grup, quan en alguna d'elles no es respecti el compromís d'individualitat grupal i/o s'utilitzin mitjans fraudulents en la seva realització, la qualificació de l'activitat serà, per a tots els membres del grup, de 0 punts (Nota Activitat=0) i sense que això exclouï la possible aplicació d'altres sancions d'acord amb el Règim Disciplinari vigent.

Qualsevol activitat no lliurada es considerarà puntuada amb zero punts

És potestatiu dels docents acceptar o no lliuraments fora dels terminis que s'indiquin. En el cas que aquests lliuraments fora de termini s'acceptin, és potestatiu del docent decidir si aplica alguna penalització i la quantia d'aquesta

## Sistema d'avaluació

---

La qualificació final és la suma ponderada de les qualificacions de les activitats d'aprenentatge:

$$Q = 0.25 \text{ PeX} + 0.25 \text{ PLab} + 0.40 \text{ Proj} + 0.10 \text{ ExiP}$$

Observacions relatives a la Recuperació

La part de teoria de l'assignatura i el Projecte (PeX i Proj) sí que és recuperable. La resta de parts no son recuperables. Per als estudiants que assisteixin a l'examen de recuperació i al lliurament projecte la seva qualificació (PeX i/o Proj) serà la obtinguda en aquesta prova i la seva qualificació final (Q) es calcularà amb les fórmules anteriorment detallades i en cap cas no serà superior a 7.

## Recursos

---

### Bàsics

Bibliografies

- An Introduction to Statistical Learning: with Applications in R 7th Edition Gareth James and Daniela Witten  
Springer Texts in Statistics – 2017  
ISBN 978-1461471370

### Complementaris

Bibliografies

- Decision Trees and Random Forests: A Visual Introduction For Beginners  
Chris Smith and Mark Koning  
Independently published (October 4, 2017)  
ISBN 978-1549893759
- Practical Machine Learning with H2O: Powerful, Scalable Techniques for Deep Learning and AI 1st Edition Darren Cook  
O'Reilly Media; 1 edition, December 2016  
ISBN 978-1491964606