

Enginyeria de Dades

Codi: 106565
Crèdits: 6

Titulació	Tipus	Curs	Semestre
2504392 Intel·ligència Artificial / Artificial Intelligence	OB	1	2

Professor/a de contacte

Nom: Javier Vazquez Corral
Correu electrònic: javier.vazquez.corral@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: anglès (eng)
Grup íntegre en anglès: Sí
Grup íntegre en català: No
Grup íntegre en espanyol: No

Prerequisits

Els estudiants hauran d'haver cursat i entès les assignatures: "Introduction to Programming I" i "Mathematical Foundations I".

Objectius

L'assignatura pretén proporcionar els fonaments de l'anàlisi i la visualització de dades. S'estudiaran les diferents etapes dels processos d'anàlisi de dades, des de la recollida, anotació i preparació de dades, fins a la seva anàlisi i visualització, preparant el camí per a una modelització més avançada mitjançant l'aprenentatge automàtic (M8).

Competències

- Conceptualitzar i modelar alternatives de solucions complexes per a problemes d'aplicació de la intel·ligència artificial en diferents àmbits, i planificar i gestionar projectes per al disseny i desenvolupament de prototips que demostrin la validesa del sistema proposat.
- Conèixer i utilitzar de manera eficient les tècniques i eines de representació, manipulació, anàlisi i gestió de dades a gran escala.
- Conèixer, comprendre, utilitzar i aplicar adequadament els fonaments matemàtics necessaris per desenvolupar sistemes de raonament, aprenentatge i manipulació de grans volums de dades.
- Introduir canvis en els mètodes i els processos de l'àmbit de coneixement per donar respostes innovadores a les necessitats i demandes de la societat.
- Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
- Treballar cooperativament per aconseguir objectius comuns, assumint la pròpia responsabilitat i respectant el rol dels diferents membres de l'equip.

Resultats d'aprenentatge

1. Concebre, dissenyar i implementar processos de recopilació i anotació de dades adequats al problema concret a resoldre.
2. Conèixer les eines bàsiques de manipulació de diferents tipus de dades estructurades, semiestructurades i no estructurades.

3. Proposar nous mètodes o solucions alternatives fonamentades.
4. Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
5. Treballar cooperativament per aconseguir objectius comuns, assumint la pròpia responsabilitat i respectant el rol dels diferents membres de l'equip.
6. Utilitzar adequadament els mètodes de visualització de dades.
7. Utilitzar mètodes d'anàlisi de dades per provar hipòtesis, i obtenir interpretacions útils.

Continguts

L'assignatura es divideix en 11 temes principals:

- Introducció.
- Repàs d'estadística bàsica.
- Numpy / Matplotlib/ Pandas
- XML/JSON
- Expressions regulars/DFA/NFA
- Tipus de dades. Dades que manquen.
- Introducció a l'anàlisi de dades. Outliers.
- PCA
- kNN
- Image retrieval i recommender systems.
- K-means

Metodologia

Hi haurà tres tipus de sessions:

Classes teòriques: L'objectiu d'aquestes sessions és que el professor expliqui el fons teòric de l'assignatura. Per a cadascun dels temes estudiats s'explica la teoria i la formulació matemàtica, així com les solucions algorítmiques corresponents.

Sessions d'exercicis: Seran sessions que faciliten la interacció. En aquestes sessions es pretén reforçar la comprensió dels temes exposats a les classes de teoria proposant casos pràctics que requereixin el disseny d'una solució en la qual s'utilitzin els mètodes vists a les classes de teoria.

Sessions pràctiques de laboratori: Seran sessions en què es realitzaran diferents tipus d'activitats relacionades amb la realització del projecte/projectes per grups d'estudiants. Durant les sessions pràctiques es presentaran el projecte/projectes a resoldre i es realitzaran una sèrie d'activitats en grups d'alumnes en modalitat de treball col·laboratiu. S'abordarà la identificació del problema, la discussió del disseny, la distribució i organització del treball a realitzar, el desenvolupament de la solució i la presentació dels resultats al professor i a la resta d'alumnes.

Tota la informació de l'assignatura i els documents relacionats que necessiten els estudiants es trobaran al campus virtual.

Dins l'horari marcat pel centre o titulació, es reservaran 15 minuts d'una classe perquè els estudiants avaluïn els seus professors i els seus cursos o mòduls mitjançant qüestionari

Nota: es reservaran 15 minuts d'una classe, dins del calendari establert pel centre/titulació, per a la complementació per part de l'alumnat de les enquestes d'avaluació de l'actuació del professorat i d'avaluació de l'assignatura/mòdul.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classes de problemes	12	0,48	2, 6, 7
Classes de teoria	28	1,12	2, 3, 7
Tipus: Supervisades			
Classes de laboratori	12	0,48	1, 2, 3, 4, 5, 6, 7
Tipus: Autònomes			
Completació dels exercicis a casa	16	0,64	
Estudi de la teoria	20	0,8	2
Realització dels projectes de pràctiques	58	2,32	1, 2, 3, 4, 5, 6, 7

Avaluació

Nota final:

La nota final es calcula mitjançant una mitjana ponderada en funció de les diferents activitats que es fan:

Nota final = $0,4 \cdot \text{Nota de teoria} + 0,2 \cdot \text{Nota d'exercicis} + 0,4 \cdot \text{Nota de laboratori}$

Per aplicar aquesta fórmula, la condició és que tant la nota de teoria com la de laboratori siguin superiors a 5.

Quan un alumne no aconseguisca el mínim exigít en alguna de les activitats d'avaluació, la nota final serà la de l'element no permetent el càlcul (és a dir, si un alumne té un 6 a Teoria, un 5a Exercicis, però un 2 a Laboratori, la nota final serà un 2).

Nota de teoria

La nota de teoria té com a objectiu avaluar les capacitats individuals de l'alumne quant als continguts teòrics de l'assignatura. Això es fa de forma continuada durant el curs a través de dos exàmens parcials:

Nota de Teoria = $0,5 \cdot \text{Nota Examen 1} + 0,5 \cdot \text{Nota Examen 2}$

L'examen 1 es fa a meitat del quadrimestre i serveix per eliminar part de la matèria si s'aprova. L'examen 2 es fa al final del quadrimestre i serveix per eliminar part de la matèria si s'aprova. Aquests exàmens tenen com a objectiu avaluar les habilitats de cada alumne de forma individualitzada, tant pel que fa a la resolució d'exercicis utilitzant les tècniques explicades a classe, com a avaluar el nivell de conceptualització que l'alumne ha realitzat de les tècniques vistes. Per obtenir una nota final de teoria aprovada, caldrà que les notes 1 i 2 de l'examen parcial siguin ambdues superiors a 4.

Examen de recuperació. En cas que la nota de teoria no assolixi el nivell adequat per aprovar, els alumnes podran fer un examen de recuperació, destinat a recuperar la part reprovada (1, 2 o ambdues) del procés d'avaluació contínua. Nota: L'examen de recuperació també es pot fer si l'alumne vol obtenir una qualificació superior a les anteriors.

Nota d'exercicis:

L'objectiu dels exercicis és que l'alumne s'entreni amb els continguts de la matèria de manera continuada i es familiaritzi amb l'aplicació dels conceptes teòrics. Com a prova d'aquest treball se sol·licita la presentació d'un portafolis on es confrontaran els exercicis realitzats:

Nota d'exercicis = Avaluació del portafoli

Nota de Pràctiques de Laboratori:

La part de pràctiques de laboratori té un pes fonamental a la nota global de l'assignatura. Les sessions de laboratori tenen com a objectiu que l'alumne dissenyi una solució a un problema plantejat de manera contextualitzada. Aquests problemes requeriran el disseny d'una solució integral, des de l'exploració de les tècniques disponibles fins al modelatge de dades. Les sessions de laboratori s'estructuren al voltant del projecte/s. Els projectes s'avaluen a través del seu lliurable, una exposició oral que els alumnes faran a classe i un procés d'autoavaluació. La qualificació es calcula de la manera següent:

Qualificació del projecte = $0,6 * \text{Qualificació dels lliurables} + 0,3 * \text{Qualificació de la presentació} + 0,1 * \text{Qualificació de l'autoavaluació}$

En cas de ser més d'un projecte, la Nota del Laboratori serà la mitjana de totes les notes dels diferents projectes. En cas de no aprovar el projecte, es permetrà la recuperació de la part dels lliurables dels projectes no aprovats, restringida a una nota màxima de 7/10. La presentació oral no es pot recuperar.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Examen parcial 1	0.2	2	0,08	1, 2, 3, 4, 7
Examen parcial 2	0.2	2	0,08	1, 2, 3, 4, 7
Portfoli d'exercicis	0.2	0	0	1, 2, 3, 4, 6, 7
Projecte de pràctiques	0.4	0	0	1, 2, 3, 4, 5, 6, 7

Bibliografia

Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media, 2015, 1st Ed.

Python Data Science Handbook, Jake Van der Plas, O'Reilly Media, 2016, 1st Ed.

Computational and Inferential Thinking: The Foundations of Data Science, Ani Adhikari and John DeNero, online: <https://ds8.gitbooks.io/textbook/content/>

Programari

Python