

Ciència de Dades

Codi: 104540
Crèdits: 6

Titulació	Tipus	Curs	Semestre
2503743 Gestió de Ciutats Intel·ligents i Sostenibles	OB	2	2

Professor/a de contacte

Nom: Dimosthenis Karatzas

Correu electrònic: Dimosthenis.Karatzas@uab.cat

Utilització d'idiomes a l'assignatura

Llengua vehicular majoritària: espanyol (spa)

Grup íntegre en anglès: No

Grup íntegre en català: No

Grup íntegre en espanyol: No

Prerequisits

Haver cursat les assignatures Informàtica, Matemàtiques, Programació d'aplicacions a Internet, de primer curs, i Bases de Dades de segon curs.

Objectius

Aquesta assignatura ha de permetre a l'estudiant conèixer les tecnologies existents i les diverses formes d'abordar la gestió i anàlisi de les dades generades per la ciutat diàriament.

Els estudiants aprendran tècniques de visualització, anàlisi i modelatge de dades que els permetran generar nou coneixement i percepcions a partir de les dades de la ciutat.

Competències

- Que els estudiants hagin demostrat que comprenen i tenen coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es basa en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda d'aquell camp d'estudi.
- Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
- Que els estudiants sàpiguen aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
- 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 destacats d'índole social, científica o ètica.
- Resoldre problemes de gestió urbana utilitzant coneixements, metodologies i procediments de disseny i implementació d'aplicacions informàtiques per a diferents tipus d'entorns (web, mòbil, núvol) i amb diferents paradigmes.
- Treballar cooperativament en entorns complexos o incerts i amb recursos limitats, en un context multidisciplinari, assumint i respectant el rol dels diferents membres de l'equip.

Resultats d'aprenentatge

1. Aplicar tècniques de presa de decisions automatitzada.
2. Que els estudiants hagin demostrat que comprenen i tenen coneixements en una àrea d'estudi que parteix de la base de l'educació secundària general, i se sol trobar a un nivell que, si bé es basa en llibres de text avançats, inclou també alguns aspectes que impliquen coneixements procedents de l'avantguarda d'aquell camp d'estudi.
3. Que els estudiants puguin transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
4. Que els estudiants sàpiguin aplicar els coneixements propis a la seva feina o vocació d'una manera professional i tinguin les competències que se solen demostrar per mitjà de l'elaboració i la defensa d'arguments i la resolució de problemes dins de la seva àrea d'estudi.
5. 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 destacats d'índole social, científica o ètica.
6. Treballar cooperativament en entorns complexos o incerts i amb recursos limitats, en un context multidisciplinari, assumint i respectant el rol dels diferents membres de l'equip.

Continguts

- Preparació de dades
 - Visualització de dades
 - Normalització
 - Valors desconeguts
 - Reducció de dimensionalitat
 - Selecció de característiques
- Classificació (tècniques supervisades)
 - Probabilitats, Classificador Naive Bayes
 - Arbres de decisió i "random forests"
 - Classificació jeràrquica
- Generació de coneixement (tècniques no supervisades)
 - Regles d'associació
 - Sistemes de recomanació

Metodologia

La ciència de dades es defineix pels tipus de problemes que intenta solucionar, per tant, serà la tipologia de problemes la que dirigirà l'Organització de tots els Continguts.

Es faran tres tipus de sessions:

Classes de teoria: Seran classes de tipus magistrals en què l'Objectiu és que el professor expliqui els Continguts teòrics de l'assignatura. Per a cada un dels temes es planteja un tipus de problema que es pretén solucionar i a partir d'aquí es van plantejant els sub-objectius i s'expliquen les corresponents solucions algorítmiques a cada un d'ells.

Classes de problemes: Seran classes que faciliten la interacció. En aquestes classes es persegueix reforçar la comprensió dels temes de la teoria plantejant casos pràctics que requereix el disseny d'una solució en la qual s'utilitzin els Mètodes vistos a les classes de teoria.

Laboratori de pràctiques: Seran sessions en què es duren a terme diferents tipus d'activitats relacionades amb la realització en equip de dos projectes de manera seqüencial. A les sessions de pràctiques es farà la presentació dels projectes a resoldre i una sèrie d'activitats que es duren a terme en equips d'estudiants en mode de treball col·laboratiu. S'haurà abordar la identificació del problema, la discussió del disseny, el repartiment i l'organització del treball a realitzar, el desenvolupament de la solució i la presentació dels resultats al professor i a la resta dels estudiants.

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

El professor donarà comentaris individualitzats per a cadascuna de les activitats lliurades pels estudiants. S'establirà un sistema de tutories i de consultes fora de l'horari de classe i es fomentarà que els estudiants facin ús de les mateixes.

La competència transversal T01 es posarà en pràctica mitjançant el treball en equip i l'intercanvi col·laboratiu que suposa el desenvolupament dels dos projectes, que s'acompanya per activitats supervisades en el laboratori de pràctiques. L'avaluació dels projectes inclou una presentació oral de cada equip a la resta de la classe, durant qual els alumnes han d'exposar el seu treball i també explicar l'organització de l'equip durant el desenvolupament del projecte.

Activitats formatives

Títol	Hores	ECTS	Resultats d'aprenentatge
Tipus: Dirigides			
Classe de Problemes	12	0,48	1, 2, 3
Classe de Teoria	26	1,04	2, 3, 4, 5
Tipus: Supervisades			
Laboratori de Pràctiques	12	0,48	1, 3, 4, 5, 6
Tutories	5	0,2	1, 3, 4, 5
Tipus: Autònomes			
Dedicació a les pràctiques	37	1,48	1, 2, 4, 5, 6
Dedicació als problemes	12	0,48	1, 2, 4, 5
Lectura i estudi de material	40	1,6	1, 2

Avaluació

Per avaluar el nivell d'aprenentatge de l'estudiant s'estableix una fórmula que combina l'aprenentatge de coneixements, la capacitat de resolució de problemes i les capacitats de treballar en equip, així com de la presentació dels resultats obtinguts.

Nota Final

La nota final es calcula ponderadament de la següent manera i d'acord amb les diferents activitats que es duen a terme:

$$\text{Nota final} = 0.3 * \text{Nota Teoria} + 0.1 * \text{Nota Problemes} + 0.6 * \text{Nota Pràctiques}$$

S'aplicarà aquesta fórmula sempre que la nota de teoria i la nota de pràctiques siguin més gran que 5. No hi ha cap restricció sobre la nota de problemes. Si fent el càlcul de la fórmula surt ≥ 5 , però no arriba al mínim exigint en alguna de les activitats d'avaluació, llavors la nota que es posarà en l'expedient serà de 4.5.

Nota de Teoria

La nota de teoria pretén valorar les capacitats individuals de l'estudiant pel que fa als continguts teòrics de l'assignatura, això es fa de forma continuada durant el curs amb dos exàmens parcials:

$$\text{Nota Teoria} = 0.5 * \text{Nota Examen 1} + 0.5 * \text{Nota Examen 2}$$

Examen 1 es fa a la meitat del semestre i serveix per eliminar part de la matèria si s'aprova.

Examen 2 es fa al final del semestre i serveix per eliminar part de la matèria si s'aprova.

Aquests exàmens pretenen fer una avaluació individualitzada de les capacitats de cada estudiant per a resoldre problemes utilitzant les tècniques explicades a classe, així com avaluar el nivell de conceptualització que l'estudiant ha fet de les tècniques vistes. Per poder tenir una nota de teoria serà necessari que les notes dels exàmens 1 i 2 siguin superiors a 4.

Examen de recuperació. En cas que la nota de teoria no arribi al nivell adequat per aprovar la teoria, els estudiants es poden presentar a un examen de recuperació, destinat a recuperar la part no superada de l'avaluació contínua.

Nota de Problemes

La part de problemes té com a objectiu aconseguir que l'estudiant s'entreni amb els continguts de l'assignatura de manera continuada i que es familiaritzi directament en l'aplicació de la teoria. Com evidència d'aquest treball es demana la presentació d'un portafoli en el qual s'aniran guardant els problemes realitzats:

$$\text{Nota Problemes} = \text{Avaluació del portfoli}$$

Nota de Pràctiques

La part de pràctiques té un pes essencial en la nota global de l'assignatura i pretén que l'estudiant s'enfronti al problema de dissenyar una solució a un problema que es planteja de manera contextualitzada i que, per tant, requereix el disseny d'una solució integral, des de l'exploració de les tècniques fins al modelatge de dades. A més, l'estudiant ha de demostrar les seves habilitats per treballar en equip i presentar convincentment els resultats.

Cadascun dels dos projectes s'avalua a través de su lliurable i una presentació oral que faran els estudiants a classe. La nota es calcula de la següent manera:

$$\text{Nota Projecte} = 2/3 * \text{Nota Lliurables} + 1/3 * \text{Nota Presentació}$$

$$\text{Nota Pràctiques} = 0.5 * \text{Nota Projecte 1} + 0.5 * \text{Nota Projecte 2}$$

En cas de no superar alguns dels projectes de pràctiques es permetrà la recuperació de la part dels lliurables (2/3 de la nota de Projecte) dels projectes no superats, però no la presentació oral.

Notes importants

Sense perjudici d'altres mesures disciplinàries que s'estimin oportunes, i d'acord amb la normativa acadèmica vigent, la comissió d'activitats irregulars per part de l'estudiant (per exemple, plagiar, copiar, deixar copiar, ...) implicarà suspendre l'activitat corresponent amb un zero (0). Les activitats d'avaluació qualificades d'aquesta forma i per aquest procediment no seran recuperables.

En cas que no es faci cap lliurament de problemes, no s'assisteixi a cap sessió de presentació dels projectes en pràctiques de laboratori i no es faci cap examen, la nota corresponent serà un "no avaluable". En un altre cas, els "no presentats" computen com un 0 per al càlcul de la mitjana ponderada.

Les matrícules d'honor es concediran als que obtinguin una nota superior o igual a 9,0 en cada part, fins al 5% dels matriculats segons ordre descendent de nota final. A criteri del professorat, també es podran concedir en altres casos.

Activitats d'avaluació

Títol	Pes	Hores	ECTS	Resultats d'aprenentatge
Exàmens	30	5	0,2	2, 3, 4
Lliurables de problemes	10	0	0	1, 2, 5
Lliurables de projectes	40	0	0	1, 2, 4, 5, 6
Presentació de projectes	20	1	0,04	2, 3, 4, 6

Bibliografia

- Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media, 2015, 1st Ed.
- Pattern Recognition and Machine Learning, Christopher Bishop, Springer, 2011
- Model-Based Machine Learning, J. Winn, C. Bishop, early access: <http://mbmlbook.com/>
- Computational and Inferential Thinking: The Foundations of Data Science, Ani Adhikari and John DeNero, online: <https://ds8.gitbooks.io/textbook/content/>