

480093 - TDS - Tractament de Dades Socioambientals

Unitat responsable:	250 - ETSECCPB - Escola Tècnica Superior d'Enginyers de Camins, Canals i Ports de Barcelona		
Unitat que imparteix:	715 - EIO - Departament d'Estadística i Investigació Operativa		
Curs:	2017		
Titulació:	MÀSTER UNIVERSITARI EN CIÈNCIA I TECNOLOGIA DE LA SOSTENIBILITAT (Pla 2013). (Unitat docent Optativa)		
Crèdits ECTS:	5	Idiomes docència:	Anglès

Professorat

Responsable: KARINA GIBERT OLIVERAS

Altres: Karina Gibert Oliveras
Miquel Sánchez-Marrè

Horari d'atenció

Horari: Envieu un mail al professor corresponent per agendar la reunió: karina.gibert@upc.edu, miquel@cs.upc.edu

Capacitats prèvies

Coneixements bàsics del paquet R
Coneixements bàsics de programació
Estadística Bàsica

Requisits

Fonaments d'Estadística Aplicada i Mesura de la Sostenibilitat i el Desenvolupament

Competències de la titulació a les quals contribueix l'assignatura

Específiques:

CE04. Aplicar adequadament, i de forma crítica i eficaç, marcs conceptuals, processos i tècniques d'obtenció i tractament de dades, estadística aplicada, modelització matemàtica, anàlisi de sistemes, sistemes d'informació geogràfica, tecnologies de la informació i les comunicacions i l'ecologia industrial a la solució de reptes de la sostenibilitat i desenvolupament sostenible.

Metodologies docents

MD1: Lecture or conference (EXP): Sharing knowledge through lectures by professors or by external guest speakers.

MD4: Tutorials of practical or theoretical works (TD): to perform an activity in the classroom, or a theoretical or practical exercise, individually or in small groups, with the advice of the teacher

and

MD6: Extensive project (PA): learning based in the design, planning and realisation in groups of a complex or extensive project or piece of work, applying and extending knowledge and writing a report on this approach and the results and conclusions

480093 - TDS - Tractament de Dades Socioambientals

Objectius d'aprenentatge de l'assignatura

L'objectiu principal d'aquesta assignatura és proporcionar una visió global de l'aplicació del Data Science a la resolució de problemes socio-ambientals reals.

Es presenta la utilització de tècniques de Minería de Dades en un procés complet de Descobriment de Coneixement, destinat a extreure informació rellevant de diferent tipus de dades socio-ambientals (enquestes, monitorització de processos, data warehouses, smart sensors...) amb l'objectiu de donar suport a la presa de decisions en fenòmens o organitzacions amb alts nivells de complexitat. El curs s'orienta a problemes socio-ambientals reals i a proporcionar els elements adients per a dissenyar eficientment i correcta els processos de Minería de Dades, sempre d'acord amb el problema real d'interès en cada aplicació. També s'analitzen les competències requerides en un Data Scientist per a tal propòsit. Es presenten els principals mètodes de minería de dades; s'entrenen alguns aspectes pràctics importants en aplicacions reals, com l'efecte d'un pre-processament incorrecte, d'una incorrecta selecció del mètode de minería de dades, d'una incorrecta interpretació de resultats, de l'assumpció d'hipòtesis falses per al procés analitzat; la comunicació efectiva de resultats als decisors o el reporting també s'analitzen. Aquestes qüestions ajuden a garantir la validesa i utilitat dels resultats finals, així com l'impacte real de l'anàlisi en el domini d'estudi. Es discutirà sobre casos socio-ambientals reals; gestió d'aigua, activitats turístiques sostenibles, pol.lució, ús del sol, s'utilitzaran per mostrar la versatilitat de la disciplina per a proporcionar un millor coneixement i millor suport a la presa de decisions en un ampli espectre de problemes socio-ambientals reals i d'alt nivell de complexitat.

Hores totals de dedicació de l'estudiantat

Dedicació total: 125h	Hores grup gran:	30h	24.00%
	Hores activitats dirigides:	15h	12.00%
	Hores aprenentatge autònom:	80h	64.00%

480093 - TDS - Tractament de Dades Socioambientals

Continguts

1. Introducció	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <p>1.1. Ciència de les Dades, Minería de dades, Descubriment de coneixement a partir de dades i Suport intel·ligent a la presa de decisions.</p> <p>1.2. Pilars de la Minería de Dades: Estadística, Intel·ligència Artificial, Sistemes d'informació, Visualització</p> <p>Activitats vinculades: Presentació del projecte a desenvolupar al llarg del curs i organització dels equips de treball</p> <p>Objectius específics:</p> <p>Es presenten les disciplines de ciència de les dades i el procés genèric de Descubriment del Coneixement a partir de dades, amb les seves etapes, incloent la de minería de dades.</p> <p>S'introdueixen els pilars disciplinaris de la minería de dades: Estadística, Intel·ligència artificial, sistemes d'informació i Visualització de dades</p> <p>Finalment, es presenta l'esquema bàsic d'un procés de descobriment del coneixement.</p>	
2. Abast, el procés de KDD.	Dedicació: 2h Grup gran: 2h
<p>Descripció:</p> <p>2.1. Tipus de problemes susceptibles de Ciència de Dades</p> <p>2.2. Dominis poc estructurats</p> <p>2.3 Coneixement a priori; Coneixement implícit. Causes i conseqüències</p> <p>2.4. Principals softwares de Minería de dades (R, weka, RapidMiner)</p> <p>Objectius específics:</p> <p>Es discuteixen problemes socio-ambientals de diferents naturaleses i llur diferent nivells de complexitat, d'acord amb la classificació proposada per Simpson. S'introdueix la gestió dels dominis poc estructurats i del coneixement implícit, causes i conseqüències.</p> <p>Es presenten algunes eines software per desenvolupar tasques de minería de dades, amb un focus especial en el sistema R.</p>	
3. Formalització del problema de Ciència de Dades i disseny d'un procés complet de Descubriment de Coneixement	Dedicació: 1h Grup gran: 1h
<p>Descripció:</p> <p>Es treballen les passes del procés de Ciència de Dades i del procés de descobriment de coneixement</p> <p>Activitats vinculades: Definir el projecte a desenvolupar, identificar les fonts de dades</p>	

480093 - TDS - Tractament de Dades Socioambientals

<p>4. Estructures de dades</p>	<p>Dedicació: 1h Grup gran: 1h</p>
<p>Descripció: 4.1 Principals fonts de dades socio-ambientals 4.2 Representació de dades i metadades</p> <p>Activitats vinculades: Construir el fitxer de metadades de les dades del projecte</p> <p>Objectius específics: S'estudien les principals estructures de dades analitzades en mineria de dades per camps socio-ambientals Importància de les metadades. Formats i continguts</p>	
<p>5. Pre-processament de dades</p>	<p>Dedicació: 5h Grup gran: 5h</p>
<p>Descripció: 5.0 Metodologia de referència per a pre-processament 5.1 Qualitat de les dades 5.2 Filtratge i mostreig 5.3. Tractament de dades mancants 5.4 Tractament de dades anòmales 5.5 Transformació de les dades i dades derivades 5.6 Feature Weighting i reducció de la dimensionalitat</p> <p>Activitats vinculades: Pre-processar les dades del projecte</p> <p>Objectius específics: Es discuteix la importància de la qualitat de les dades i les conseqüències de la manca de qualitat. S'introdueixen els aspectes més rellevants de la fase de preparació de dades: Dades mancants, detecció de dades anòmales, variables derivades, transformacions, filtratge, mostreig, selecció i rellevància de les variables, reducció de la dimensionalitat (selecció de variables i mètodes factorials), tots crítics per a garantir la validesa del procés d'anàlisi. Es proporcionaran guies de bones pràctiques, així com una metodologia general de referència</p>	

480093 - TDS - Tractament de Dades Socioambientals

6. Selecció del mètode de mineria de dades	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <ul style="list-style-type: none"> 6.1. L'aproximació orientada al problema 6.2. Criteris que determinen el mètode de mineria de dades adient 6.3. El mapa conceptual de mètodes de mineria de dades (mapa DMMDM) <p>Objectius específics:</p> <p>El curs segueix un enfoc orientat al problema, on la naturalesa intrínseca del problema determina el procés d'anàlisi i no al revés. Es discuteix sobre els factors que determinen la correcta elecció del mètode de mineria de dades a emprar en un procés concret de Ciència de Dades. El mapa DMMDM presenta una tipologia de mètodes de mineria de dades que s'aporta com a base conceptual per a la selecció.</p>	
7. Fase de mineria de dades: Mètodes descriptius	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <ul style="list-style-type: none"> 7.1. Mètodes descriptius clustering: Mètodes de particions, jeràrquics. Escalabilitat. Mètodes híbrids, introducció de coneixement expert a priori. Explicitació de coneixement 7.2 Caracterització de les classes <p>Activitats vinculades:</p> <p>Fer clustering de les dades del projecte</p> <p>Objectius específics:</p> <p>S'estudien mètodes d'identificació i caracterització de perfils</p>	
8. Fase de Mineria de dades: Mètodes associatius	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <ul style="list-style-type: none"> 8.1. Regles d'associació 8.2. Mètodes factorials 8.3. Xarxes baessianes <p>Activitats vinculades:</p> <p>Aplicar algun mètode associatiu a les dades del projecte</p> <p>Objectius específics:</p> <p>Aquest tema introdueix mètodes per a descobrir les relacions d'associació entre les variables de la base de dades</p>	

480093 - TDS - Tractament de Dades Socioambientals

<p>9, Fase de mineria de dades: Mètodes discriminants</p>	<p>Dedicació: 2h 30m Grup gran: 2h 30m</p>
<p>Descripció: 9.1 Arbres de decisió 9.2 Inducció de regles de classificació 9.3 Màquines de vector de suport 9.4 anàlisi discriminant 9.5 mètodes d'ensemble i bagging 9.6 mètodes híbrids</p> <p>Activitats vinculades: Aplicar almenys dos mètodes discriminants a les dades del projecte</p> <p>Objectius específics: Mètodes per predir una variable de classe o una variable qualitativa. Se'n veuran almenys 3.</p>	
<p>10. Fase de mineria de dades: Mètodes predictius</p>	<p>Dedicació: 2h 30m Grup gran: 2h 30m</p>
<p>Descripció: 10.1 Regressió, model.lització estadística en general 10.2 Mètodes temporals 10.3 Xarxes neuronals artificials 10.4 Swarm intelligence</p> <p>Activitats vinculades: Predicció d'una o varies variables numèriques</p> <p>Objectius específics: Mètodes per predir una variable numèrica. Se'n veuran almenys 2</p>	
<p>11. Minería de dades espai-temporal</p>	<p>Dedicació: 2h 30m Grup gran: 2h 30m</p>
<p>Descripció: 11. Models espaitemporals</p> <p>Activitats vinculades: Revisió general del projecte</p> <p>Objectius específics: s'introdueixen eines per tractar dades espai-temporals</p>	

480093 - TDS - Tractament de Dades Socioambientals

12. Post-processament i validació	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <ul style="list-style-type: none"> 12.1 Eines de post-processament 12.2. Validació de models 12.3. Validació de resultats <p>Activitats vinculades:</p> <ul style="list-style-type: none"> Validació dels models del projecte <p>Objectius específics:</p> <ul style="list-style-type: none"> Eines de post-processament i validació per models i resultats adaptades als diferents mètodes de mineria de dades. 	

13. Reporting i comunicació de resultats	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <ul style="list-style-type: none"> 13.1. Informes, informes automàtics 13.2. Comunicació de resultats <p>Activitats vinculades:</p> <ul style="list-style-type: none"> Revisió de l'informe del projecte <p>Objectius específics:</p> <ul style="list-style-type: none"> Crucial per garantir que els resultats del procés de ciència de dades aportin suport efectiu a la presa de decisions i que l'anàlisi tingui impacte real en el domini objecte d'interès 	

Planificació d'activitats

Presentació intermèdia de projectes	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <ul style="list-style-type: none"> Presentació oral de la primera part del projecte i discussió Entregable escrit 	
Presentació final de projectes	Dedicació: 2h 30m Grup gran: 2h 30m
<p>Descripció:</p> <ul style="list-style-type: none"> Presentació oral i entrega escrita del projecte complet. Discussió general i individual amb els professors 	

480093 - TDS - Tractament de Dades Socioambientals

Sistema de qualificació

Es desenvoluparà un projecte d'envergadura per grups, aplicant un procés complet de data science a dades reals, el qual inclourà l'aplicació de mètodes vistos a classe. El projecte es realitza sota la supervisió del professor.

Una entrega intermèdia (D1) ajudarà a planificar millor el treball global (D2). Hi haurà una nota que s'assignarà de la següent forma:

$$\text{NotaFinal} = 0.4D1 + 0.6D2$$

on

$D1 = 0.4 * \text{qualitat del document escrit} + 0.3 * \text{qualitat de la presentació oral i discussió} + 0.2 * \text{rendiment individual mostrat a les sessions de laboratori}$

$D2 = \text{alfa} * (0.4 * \text{qualitat del document escrit} + 0.3 * \text{qualitat de la presentació oral i discussió} + 0.2 * \text{rendiment individual mostrat a les sessions de laboratori})$

essent alfa un factor entre 0.5 i 1.5 que resulta del procés de cross-evaluation realitzat pels companys de grup de treball a data d'entrega del treball final.

AV2. Prova oral de control de coneixements (PO).

AV3. Treball realitzat en forma individual o en grup al llarg del curs (TR). Inclou tant l'avaluació de resultats i informes, com la presentació oral dels mateixos.

AV4. Assistència i participació en classes i laboratoris (AP).

AV5. Qualitat i rendiment del treball en grup (TG).

Bibliografia