

**FITXA IDENTIFICATIVA****Dades de l'Assignatura**

Codi	36516
Nom	Dades no estructurades
Cicle	Grau
Crèdits ECTS	6.0
Curs acadèmic	2023 - 2024

Titulació/titulacions

Titulació	Centre	Curs	Període
1332 - Grau en Intel·ligència i Analítica de Negocis/BIA	Facultat d'Economia	4	Primer quadrimestre

Matèries

Titulació	Matèria	Caràcter
1332 - Grau en Intel·ligència i Analítica de Negocis/BIA	24 - Eines i Tècniques d'Anàlisi de Dades	Obligatòria

Coordinació

Nom	Departament
MONTORO PONS, JUAN DE DIOS	110 - Economia Aplicada
MORILLAS JURADO, FRANCISCO GABRIEL	110 - Economia Aplicada

RESUM

Amb el desenvolupament de la Web 2.0 i les Xarxes Socials van arribar les dades massives. La majoria d'aquestes dades sense una estructura inicial prefixada. Ens referim a documents, fotos, vídeos, correus electrònics, tuits, etc. La major part d'aquestes dades els generen els propis usuaris. Tots els dies s'envien correus electrònics, es publiquen tuits o es pugen fotos o vídeos a les diferents xarxes socials. Totes aquestes dades faciliten molta informació sobre les persones que l'estan generat. No obstant això, en ser dades No estructurades s'haurà de transformar la informació que generen per a poder analitzar-la i convertir-la en coneixement. Un exemple és la informació d'una empresa en la xarxa social Twitter. Es poden recopilar tots els tuits que esmenten una empresa, i mitjançant una transformació en una base de dades estructurada es podrà fer anàlisi de sentiment i saber què opinen les persones d'aquesta empresa. També es poden descarregar els tuits de la competència per a saber l'opinió sobre aqueixes empreses i poder modificar l'estratègia de l'empresa. Les dades No estructurades precisaran de bases de dades no relacionals, com poden ser MongoDB (basada en documents), Neo4j (basada en grafs), etc. Les noves bases de dades permetran realitzar una anàlisi diferent de la què es realitza amb les bases de dades relacionals. L'objectiu de l'assignatura serà que els alumnes coneguen i puguin treballar amb aquestes



dades no estructurades i realitzar anàlisis d'aquestes per a tindre un millor coneixement de l'empresa i de la competència.

CONEIXEMENTS PREVIS

Relació amb altres assignatures de la mateixa titulació

No heu especificat les restriccions de matrícula amb altres assignatures del pla d'estudis.

Altres tipus de requisits

És recomanable tindre coneixements del programari R i RStudio, així com Estadística Bàsica, Inferència Estadística i Aprenentatge Automàtic (machine Learning).

COMPETÈNCIES

1332 - Grau en Intel·ligència i Analítica de Negocis/BIA

- Que els estudiants puguen transmetre informació, idees, problemes i solucions a un públic tant especialitzat com no especialitzat.
- Capacitat per a analitzar i buscar informació provinent de fonts diverses.
- Capacitat d'aprenentatge autònom.
- Capacitat per a utilitzar les TIC, tant en l'àmbit d'estudi com en el desenvolupament professional.
- Capacitat per a definir, resoldre i exposar de forma sistèmica problemes complexos.
- Capacitat per a treballar en equip, amb el compromís per la qualitat, l'ètica, la igualtat entre persones i la responsabilitat social.
- Aplicar mètodes i tècniques d'anàlisis, síntesis i representació gràfica mitjançant programes informàtics.
- Reorganitzar i reestructurar variables i bases de dades.
- Aplicar mineria de dades mitjançant programes informàtics.
- Aplicar, utilitzant programari, tècniques d'aprenentatge automàtic (machine learning) supervisat.
- Aplicar, utilitzant programari, tècniques d'aprenentatge automàtic (machine learning) no supervisat i semisupervisat.
- Manejar aplicacions de descàrrega, capturar i manipular valors de dades no estructurades.

RESULTATS DE L'APRENTATGE

- Sintetitzar informació.



- Saber dissenyar i implementar una investigació i la seua aplicació a l'entorn econòmic.
- Aplicar les tècniques d'aprenentatge automàtic a problemes de negoci.
- Conèixer els diferents tipus de bases de dades existents i els tipus de dades que utilitzen.
- Aprendre els conceptes de la Teoria de Grafs.
- Conèixer com funciona un sistema de recomanació.
- Aprendre a processar dades no numèriques mitjançant programes informàtics.
- Aprendre a realitzar diccionaris amb els quals realitzar processament automàtic de llenguatge natural per a la seua explotació de negoci.
- Saber implementar una investigació i la seua aplicació a l'entorn econòmic amb dades textuais.
- Manejar eines quantitatives avançades i la seua aplicació a l'entorn de negocis.
- Aprendre a manejar APIs de descàrrega de dades en xarxes socials.
- Aprendre a reconèixer patrons en dades en format text.
- Aprendre processos per a generar recomanacions de compra.

DESCRIPCIÓ DE CONTINGUTS

1. Introducció

- a. Dades, Informació i Aprofitament
- b. Dades estructurades vs no estructurades
- c. La il·lusió de l'anàlisi de dades no estructurades
- d. Estàndards de marcatge: XML, JSON,...

2. Captura de Dades a la Web I

- a. Captura de dades de llocs web: captura directa
- b. Una aplicació a obtenir un índex: IPC
- c. Captura de dades mitjançant API's
- d. Bones pràctiques

3. Web Scraping: Captura de Dades a la Web II

- a. Estructura d'una pàgina web: HTML
- b. HTML dinàmic: JavaScript
- c. Fulles d'estil: CSS, W3-CSS, Bootstrap(CSS)
- d. Captura completa d'un lloc web
- e. Captura específica de dades de la web



f. Scripts que imiten la intervenció humana: Bots

4. Mineria de Text

- a. Tipus i usos de les tècniques d'anàlisi de text
- b. Enfocaments: bag of words (BoW) i natural language processing (NLP)
- c. El flux de treball bàsic
- d. Preparació i estructures de dades
- e. Visualitzacions i anàlisi descriptiu

5. Anàlisi de Sentiments

- a. Contingut subjectiu en textos, polaritat i emocions
- b. Enfocament basat en diccionaris vs aprenentatge automàtic
- c. Diccionaris de sentiments en Espanyol
- d. Aplicacions de l'anàlisi de sentiments

6. Aprenentatge Automàtic i Analítica de Text

- a. Tècniques d'aprenentatge supervisat i no supervisat
- b. Modelització probabilística de temes (topic modeling)
- c. Classificació de textos
- d. Modelització predictiva
- e. Tècniques avançades per a l'anàlisi de text

7. Sistemes de Recomanació

- a. Tipus de sistemes de recomanació
- b. Preferències implícites i explícites
- c. Construcció d'un sistema de recomanació
- d. Avaluació d'un sistema de recomanació
- e. Aplicacions

8. Xarxes i Xarxes Socials

- a. Definició i Tipologies
- b. Models estadístics de xarxes socials
- c. Models de grafs aleatoris
- d. Xarxes socials i segmentació
- e. Xarxes neuronals aplicades

**VOLUM DE TREBALL**

ACTIVITAT	Hores	% Presencial
Pràctiques en aula informàtica	45,00	100
Classes de teoria	15,00	100
Elaboració de treballs en grup	10,00	0
Elaboració de treballs individuals	10,00	0
Preparació d'activitats d'avaluació	10,00	0
Preparació de classes de teoria	20,00	0
Preparació de classes pràctiques i de problemes	30,00	0
Resolució de casos pràctics	10,00	0
TOTAL	150,00	

METODOLOGIA DOCENT

- Classe teòrica presencial per a presentar els continguts teòrics essencials de la matèria.
- Classes pràctiques presencials, relatives a la resolució de problemes, estudis de casos, amb aplicació de tècniques, presentacions orals, debats, individualment i/o en equip.
- Treball autònom supervisat i basat en la realització d'exercicis, casos pràctics i qüestions a debatre o experiments en línia, amb suport tutorial.

AVALUACIÓ

Avaluació consistent en una o diverses proves que consideraran tant qüestions teòriques com pràctiques. (50%)

Avaluació de les activitats pràctiques desenvolupades per el/la alumne/a durant el curs, a partir de l'elaboració de treballs/memòries i/o exposicions orals, amb defensa de les posicions desenvolupades per el/la alumne/a. (30%)

Avaluació contínua de cada alumne, basada en la participació i grau d'implicació del/de l'alumne/a en el procés d'ensenyament-aprenentatge, tenint en compte l'assistència regular a les activitats presencials previstes i la resolució de qüestions i problemes proposats periòdicament. (20%)

REFERÈNCIES



Bàsiques

- Anandarajan, M.; Hill, C. y Nolan, T. (2019) Practical Text Analytics. Maximixing the Value of Text Data. Springer Nature, Switzerland.
- Bali, R. Sarkar, D. y Sharma, T. (2017) Learning Social Media Analytics with R. Pack Publishing, Birmingham, UK.
- Gorakala, S.K. y Usuelli, M. (2015) Building a Recommendation System with R. Pack Publishing, Birmingham, UK.
- Kwartler, T. (2017). Text mining in practice with R. John Wiley & Sons.
- Munzert, S.; Rubba, C.; Meiner, P. y Nyhuis, D. (2015) Automated data collection with R: a practical guide to web scraping and text mining. John Wiley & Sons, UK.
- Rusell, Matthew A. (2013) Mining the Social Web. O'Reilly Media, CA.