



Guía docente

Identificación de la asignatura

Asignatura / Grupo	11651 - Minería de Textos para las Ciencias Sociales / 1
Titulación	Máster Universitario en Análisis de Datos Masivos en Economía y Empresa
Créditos	3
Período de impartición	Segundo semestre
Idioma de impartición	Castellano

Profesores

Horario de atención a los alumnos

Profesor/a	Hora de inicio	Hora de fin	Día	Fecha inicial	Fecha final	Despacho / Edificio
Antonio Bibiloni Coll toni.bibiloni@uib.es	Hay que concertar cita previa con el/la profesor/a para hacer una tutoría					

Contextualización

La asignatura pretende proporcionar al estudiante una perspectiva general de las técnicas y herramientas más usuales para extraer conocimiento a partir de información no estructurada procedente de textos (tales como páginas web, redes sociales, informes, etc.). Se estudiarán todo un conjunto de herramientas relacionadas con la temática

Requisitos

Esta asignatura no tiene requisitos específicos. Es recomendable haber cursado previamente las asignaturas "Tecnologías para el análisis de datos masivos" y "Aplicaciones de la minería de datos a la industria del turismo".

Competencias

Específicas

- * CE1 - Capacidad para procesar grandes volúmenes documentales para extraer patrones y conocimiento mediante técnicas de minería de textos y minería web
- * CE10 - Conocer los ámbitos de aplicación del paradigma "Big Data" y desarrollar la capacidad para extender el análisis de datos a actividades estratégicas en economía, empresa y turismo
- * CE14 - Diseñar y planificar un proyecto de análisis de datos en problemas reales del ámbito económico, empresarial o turístico



Guía docente

Genéricas

- * CG1 - Saber recuperar datos y extraer conocimiento de grandes volúmenes de datos mediante la aplicación eficiente de técnicas de análisis de datos en diferentes dominios. Adoptar los modos de interacción adecuados según las tareas de usuario que se estén apoyando, en especial en aquellos casos en los que interviene el razonamiento analítico
- * CG3 - Conocer aplicaciones avanzadas de la ciencia de datos y de sus tecnologías a la economía, empresa y turismo
- * CG6 - Capacidad para aplicar los principios de la ciencia de datos a la economía, la empresa y el turismo, así como conocer la legislación, regulación y normalización asociada al uso de datos

Básicas

- * Se pueden consultar las competencias básicas que el estudiante tiene que haber adquirido al finalizar el máster en la siguiente dirección: http://estudis.uib.cat/es/master/comp_basiques/

Contenidos

Contenidos temáticos

Minería de datos. Minería de textos para las ciencias sociales
Introducción al acceso, captura y organización de los datos.

Machine Learning aplicado a la minería de textos.

Minería de datos versus minería de textos

Análisis. Técnicas

Explicación teórica de los modelos más utilizados en el análisis de textos.

Clasificación de documentos.

Text Analytics

Herramientas disponibles para su uso.

Sentimiento. Análisis de opiniones

Análisis de opiniones en textos.

Análisis de sentimientos y/o sensaciones.

Natural Language Processing.

Herramientas disponibles para su uso.

ChatBot. NLP

Concepto de NLP aplicado.

Ejemplo de Chatbot aplicado.

Metodología docente

El contenido teórico se expondrá en clases presenciales de teoría basadas en textos de referencia a los que el alumno tendrá acceso a través de la biblioteca. Los conceptos teóricos presentados serán aplicados a la



Guía docente

resolución de casos de estudio, tanto durante las clases expositivas (cuando sea apropiado), como en clases específicas de prácticas, o en tutorías en grupo reducido o individual.

El alumno resolverá prácticas sencillas de refuerzo de los conceptos y técnicas vistas en clase. Asimismo, para profundizar en dichas técnicas, se propondrán casos de estudio de complejidad superior. El seguimiento de dicho trabajo se realizará tanto en clase como a través de tutorías, donde en grupo reducido o a nivel individual se procederá a la discusión e intercambio de información entre alumno(s) y profesor. Este tipo de actividad puede llevar asociada la exposición oral de trabajos por parte de los alumnos.

Con el propósito de favorecer la autonomía y el trabajo personal del alumno, la asignatura forma parte del proyecto Campus Extens. Este proyecto incorpora el uso de herramientas telemáticas para conseguir una enseñanza universitaria flexible y a distancia. De esta forma, el alumno dispondrá de documentos electrónicos y enlaces a Internet relacionados con los contenidos de la asignatura, enunciados de problemas/ prácticas.

Actividades de trabajo presencial (0,72 créditos, 18 horas)

Modalidad	Nombre	Tip. agr.	Descripción	Horas
Clases teóricas	Clases con exposición de contenidos	Grupo grande (G)	Esta metodología se centra en la exposición verbal por parte del profesor de los contenidos sobre la materia objeto de estudio. El profesor suministra a los alumnos información esencial y organizada procedente de diversas fuentes con unos objetivos específicos predefinidos. En la clase magistral el profesor debe motivar a los alumnos, exponer los contenidos de un tema, explicar conocimientos, efectuar demostraciones si procede, presentar experiencias, etc. El uso de las nuevas tecnologías, así como el uso de recursos audiovisuales, se hace imprescindible actualmente en el método expositivo. El profesor debe fomentar también la participación del alumno en las clases como un agente activo y no meramente pasivo. En el uso de este tipo de metodología el profesor utilizará todos los recursos educativos necesarios: proyección de diapositivas y otros documentos interactivos, vídeos, etc.	9
Clases prácticas	Estudio de casos	Grupo grande (G)	El proceso consiste en la presentación por parte del profesor de un caso concreto, de extensión variable según el diseño organizativo, para su estudio junto con un guión de trabajo que oriente dicho proceso. Así esta metodología lleva implícita una parte analítica con la finalidad de conocer el problema, interpretarlo, resolverlo y contemplar distintos escenarios y procedimientos alternativos de solución. Se suele desarrollar en tres etapas, una primera de presentación y familiarización del tema, una segunda de análisis detenido del caso y una tercera de preparación de conclusiones y recomendaciones. Dependiendo de la extensión temporal de los casos de estudio, la incorporación de los recursos de documentación e información on-line puede resultar particularmente interesante. Las TICs permiten que los integrantes del grupo de estudio y el profesor estén en contacto permanente y extender temporalmente todos los procesos interactivos.	9

Al inicio del semestre estará a disposición de los estudiantes el cronograma de la asignatura a través de la plataforma UIBdigital. Este cronograma incluirá al menos las fechas en las que se realizarán las pruebas de evaluación continua y las fechas de entrega de los trabajos. Asimismo, el profesor o la profesora informará a los estudiantes si el plan de trabajo de la asignatura se realizará a través del cronograma o mediante otra vía, incluida la plataforma Aula Digital.

Guía docente

Actividades de trabajo no presencial (2,28 créditos, 57 horas)

Modalidad	Nombre	Descripción	Horas
Estudio y trabajo autónomo individual	Estudio y trabajo individual	Preparación de seminarios, lecturas, investigaciones, trabajos, memorias, obtención y análisis de datos, etc. para exponer o entregar en clase mediante el trabajo de los alumnos de forma individual. Incluye además, el estudio personal (preparar exámenes, trabajo en biblioteca, lecturas complementarias, hacer problemas y ejercicios, etc.), que es fundamental para el aprendizaje autónomo.	57

Riesgos específicos y medidas de protección

Las actividades de aprendizaje de esta asignatura no conllevan riesgos específicos para la seguridad y salud de los alumnos y, por tanto, no es necesario adoptar medidas de protección especiales.

Evaluación del aprendizaje del estudiante

Fraude en elementos de evaluación

De acuerdo con el artículo 33 del Reglamento Académico, "con independencia del procedimiento disciplinario que se pueda seguir contra el estudiante infractor, la realización demostrablemente fraudulenta de alguno de los elementos de evaluación incluidos en guías docentes de las asignaturas comportará, a criterio del profesor, una minusvaloración en su calificación que puede suponer la calificación de «suspense 0» en la evaluación anual de la asignatura".

Estudio de casos

Modalidad	Clases prácticas
Técnica	Trabajos y proyectos (no recuperable)
Descripción	El proceso consiste en la presentación por parte del profesor de un caso concreto, de extensión variable según el diseño organizativo, para su estudio junto con un guión de trabajo que oriente dicho proceso. Así esta metodología lleva implícita una parte analítica con la finalidad de conocer el problema, interpretarlo, resolverlo y contemplar distintos escenarios y procedimientos alternativos de solución. Se suele desarrollar en tres etapas, una primera de presentación y familiarización del tema, una segunda de análisis detenido del caso y una tercera de preparación de conclusiones y recomendaciones. Dependiendo de la extensión temporal de los casos de estudio, la incorporación de los recursos de documentación e información on-line puede resultar particularmente interesante. Las TICs permiten que los integrantes del grupo de estudio y el profesor estén en contacto permanente y extender temporalmente todos los procesos interactivos.
Criterios de evaluación	Exponer, participar, debatir, intercambiar opiniones con el profesor y el resto de alumnos con el fin de adquirir las competencias de la asignatura

Porcentaje de la calificación final: 40% con calificación mínima 5

Guía docente

Estudio y trabajo individual

Modalidad	Estudio y trabajo autónomo individual
Técnica	Informes o memorias de prácticas (no recuperable)
Descripción	Preparación de seminarios, lecturas, investigaciones, trabajos, memorias, obtención y análisis de datos, etc. para exponer o entregar en clase mediante el trabajo de los alumnos de forma individual. Incluye además, el estudio personal (preparar exámenes, trabajo en biblioteca, lecturas complementarias, hacer problemas y ejercicios, etc.), que es fundamental para el aprendizaje autónomo.
Criterios de evaluación	A través de la presentación de informes de los casos de estudio se evaluará el nivel de adquisición de los contenidos y las competencias específicas de la asignatura. Cada caso podrá tener una parte consistente en la resolución de problemas y siempre contendrá una parte relacionada con los conceptos teóricos.

Porcentaje de la calificación final: 60% con calificación mínima 5

Recursos, bibliografía y documentación complementaria

Bibliografía básica

- * **Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications.** Miner, Gary. Academic Press, , 2012. (ISBN: 978-0-12-386979-1).
- * **R and Data Mining: Examples and Case Studies.** Yanchang Zhao. Academic Press, 2013. (ISBN:9780123969637).

Otros recursos

- * Artículos científicos y divulgativos de los contenidos del curso.
- Choi, H., & Varian, H. (2012). *Predicting the present with Google Trends*. Economic Record, 88(s1), 2-9.
- Duan, W., Yu, Y., Cao, Q., & Levy, S. (2015). *Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach*. Cornell Hospitality Quarterly.
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). *What can big data and text analytics tell us about hotel guest experience and satisfaction?* International Journal of Hospitality Management, 44, 120-130.
- Mostafa, M. M. (2013). *More than words: Social networks' text mining for consumer brand sentiments*. Expert Systems with Applications, 40(10), 4241-4251.