19-20

MÁSTER UNIVERSITARIO EN
INGENIERÍA Y CIENCIA DE DATOS

GUÍA DE ESTUDIO PÚBLICA



MINERÍA DE DATOS DE LOS MEDIOS SOCIALES

CÓDIGO 31110107

UNED

19-20

MINERÍA DE DATOS DE LOS MEDIOS SOCIALES

CÓDIGO 31110107

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

EQUIPO DOCENTE

HORARIO DE ATENCIÓN AL ESTUDIANTE

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

RESULTADOS DE APRENDIZAJE

CONTENIDOS

METODOLOGÍA

SISTEMA DE EVALUACIÓN

BIBLIOGRAFÍA BÁSICA

BIBLIOGRAFÍA COMPLEMENTARIA

RECURSOS DE APOYO Y WEBGRAFÍA

ADENDA AL SISTEMA DE EVALUACIÓN CON MOTIVO DE LA PANDEMIA COVID 19

Nombre de la asignatura	MINERÍA DE DATOS DE LOS MEDIOS SOCIALES
Código	31110107
Curso académico	2019/2020
Título en que se imparte	MÁSTER UNIVERSITARIO EN INGENIERÍA Y CIENCIA DE DATOS
Tipo	CONTENIDOS
Nº ETCS	4
Horas	100.0
Periodo	SEMESTRE 2
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

El objetivo principal de la asignatura Minería de datos de los medios sociales es conseguir que los alumnos del Master Universitario en Ciencia de Datos de la Escuela Superior de Ingeniería Informática de la UNED sean capaces de adquirir el conocimiento y habilidades para analizar entornos sociales en la web. La asignatura se centra en el análisis de las relaciones de los actores involucrados en entornos dinámicos y complejos como son la web social.

La web social o los medios sociales en la web son aquellos entornos que hacen posible la relación entre personas por medio de la web. Las redes sociales son un ejemplo pero no es el único. El extenso uso que se hace de la web social las convierten en una potente e importante herramienta en la sociedad actual.

Para realizar análisis en la web social utilizaremos una metodología heredera de la Minería de datos. Primero, hay que extraer datos de interés, filtrarlos y prepararlos. Este primer punto será estudiado en el Temas 1. Segundo, se enseñará el análisis de redes sociales (SNA en su siglas en inglés) como medio útil de análisis y visualización de los datos debido a entornos sociales virtuales. Este punto será estudiado en el Tema 2. El SNA ha dado un salto tecnológico importante al tener en cuenta el dinamismo de las redes, es decir, las relaciones entre los nodos de una red evolucionan con el tiempo. Se introducirá el dinamismo de las redes en el Tema 3. Por último, se enseñarán técnicas de Aprendizaje Automático que sirven para extraer conocimiento de datos, completando así el análisis de redes sociales. Este último se estudiará en el Tema 4.

El entorno virtual será la herramienta más importante para la gestión de esta asignatura. Al ser un entorno de aprendizaje colaborativo, se animará y estimulará a que los estudiantes compartan dudas, comentarios, ideas y proyectos con el resto. De esta forma se fomentará las habilidades de diálogo y negociación, tan útiles en entornos de trabajo reales.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Los estudiantes de esta asignatura no requieren conocimientos previos adicionales más allá de los que surgen del seguimiento de los requeridos por el máster, permitiendo utilizar y desarrollar las competencias correspondientes.

Es conveniente dominar el inglés técnico (leer y escribir) para manejar con facilidad las fuentes bibliográficas.

EQUIPO DOCENTE

Nombre y Apellidos	ANTONIO RODRIGUEZ ANAYA (Coordinador de asignatura)
Correo Electrónico	arodriguez@dia.uned.es
Teléfono	91398-6550
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

Nombre y Apellidos	ANGELES MANJARRES RIESCO
Correo Electrónico	amanja@dia.uned.es
Teléfono	91398-8125
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	INTELIGENCIA ARTIFICIAL

HORARIO DE ATENCIÓN AL ESTUDIANTE

La tutorización de los estudiantes tendrá lugar esencialmente a través de los foros de la plataforma, aunque también podrán utilizarse ocasionalmente otros medios, tales como chats interactivos, servicios de mensajería instantánea y el correo electrónico. Adicionalmente está también previsto, para temas personales que no afecten al resto de los estudiantes, atender consultas en persona o por teléfono. Se aconseja, en cualquier caso, el uso del correo electrónico para contactar con el equipo docente.
Los datos de contacto del equipo docente son los siguientes:
Antonio Rodríguez Anaya - Despacho 3.04 (E. T. S. I. Informática) Tel.: 91 398 65 50 e-mail: arodriguez@dia.uned.es
Manuel Luque Gallego - Despacho 3.05 (E. T. S. I. Informática) Tel.: 91 398 84 05 e-mail: mluque@dia.uned.es
El horario de guardia del equipo docente será los lunes de 15:00 a 19:00.
Horario de atención al estudiante: Martes de 10:00 a 14:00 y de 15:00 a 19:00.

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

COMPETENCIAS BÁSICAS
CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades. sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.
CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y

sin ambigüedades.
CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

COMPETENCIAS GENERALES

CG1 - Identificar los métodos apropiados para la solución de problemas asociados a la ciencia de datos y la analítica de información
CG2 - Ser capaz de aplicar diferentes técnicas de aprendizaje máquina, seleccionando el algoritmo óptimo que genere modelos precisos y permita el desarrollo de soluciones predictivas en diferentes ámbitos de uso

COMPETENCIAS TRANSVERSALES

CT1 - Ser capaz de abordar y desarrollar proyectos innovadores en entornos científicos, tecnológicos y multidisciplinares.
CT2 - Ser capaz de tomar decisiones y formular juicios basados en criterios objetivos (datos experimentales, científicos o de simulación disponibles).

COMPETENCIAS ESPECÍFICAS

CE4 - Conocer las diferentes librerías e implementaciones de las técnicas de representación de información, especialmente relevantes en la visualización de datos masivos
CE5 - Desarrollar modelos de aprendizaje máquina (Machine Learning) basados en las diferentes categorías de clasificación: supervisada, no supervisada y semi-supervisada

RESULTADOS DE APRENDIZAJE

Los resultados más relevantes que se pretenden alcanzar con el estudio de esta asignatura son los siguientes:

•Aplicar técnicas de extracción de información de una red social.
•Conocer herramientas de análisis de redes sociales.
•Desarrollar un análisis de la red social.
•Proponer la visualización de los resultados para facilitar su comprensión e interpretación
•Contrastar las diferencias entre un análisis dinámico y un análisis estático de una red social.
•Identificar distintas técnicas de aprendizaje automático con las que completar el proceso de minería de datos en la web social.

CONTENIDOS

Tema 1: Herramientas y métodos de extracción de información en redes sociales

El objetivo de este tema es familiarizarse con aplicaciones y ayudas para obtener información relevante de las interacciones que se realizan en la Web social. La Web social es amplia, muy amplia, pero hay sitios web que llevan tiempo teniendo una gran importancia. Estamos hablando de las redes sociales y de dos fundamentalmente: Facebook, la más popular actualmente; Twitter, la red social de mayor relevancia informativa.

Sin embargo, los contenidos del tema no se restringen exclusivamente a dos redes sociales, sino a la información relevante que se puede extraer. Es interesante, por tanto, tener nociones de teorías sociales, las cuales explican el comportamiento humanos en grupos. Al realizar una tarea de minería de datos sobre los datos de la web social, se tendrá en mente un objetivo: ¿qué se quiere deducir?, ¿qué se está buscando? Las teorías sociales ofrecen un contexto teórico para encauzar el proceso de minería.

Este tema tiene un carácter principalmente práctico y nos centraremos en la obtención de información relevante utilizando las ayudas que las redes sociales suministran u otras aplicaciones de autor que son ajenas al propio desarrollo de las redes sociales. Para ello se propone un trabajo práctico, una tarea individual obligatoria.

El equipo docente ofrecerá la documentación necesaria para poder utilizar las ayudas y aplicaciones que son necesarias para obtener la información relevante.

Tema 2: Análisis y visualización de redes sociales estáticas y dinámicas

En este tema nos centraremos en el análisis de redes sociales (SNA en sus siglas en inglés). El análisis de redes es el área encargada de estudiar las redes mediante la teoría de grafos. Las redes pueden ser de diversos tipos: social, transporte, eléctrica, biológica, internet, información, epidemiología, etc. Por eso mismo, el SNA no se restringe únicamente a redes sociales entendidas como sitios web de interacción social como Facebook o Twitter, sino a todo contexto en el que se puede formar una red de interacciones.

En este tema se da una introducción a los fundamentos teóricos del SNA pero el objetivo es comenzar a utilizar herramientas que nos hagan posible hacer SNA. Además, gracias al tema anterior, ya se sabe obtener información de sitios web en donde las interacciones se organizan como redes. Las comunicaciones en una red social, el foro del curso virtual, son ejemplos claros de redes a los que poder aplicar un SNA.

La web social es un entorno dinámico. Constantemente ocurren nuevas interacciones, se envían mensajes, se comentan publicaciones, o se borran. Sin embargo, la visión común de SNA tiene un carácter de tomar una fotografía de un momento dado de una red. En este tema queremos introducir el dinamismo en el análisis y entender la evolución dinámica de los atributos de las redes formadas en la web social.

Este tema tiene un carácter principalmente práctico ya que lo importante es familiarizarse con las posibilidades que ofrece un SNA para representar y analizar lo que está sucediendo en una red.

Tema 3: Técnicas de minería de datos para redes sociales

Para estudiar este tema hay que conocer un área de las ciencias de la computación muy importante, la minería de datos. La minería de datos o exploración de datos es un proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los

métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.

Hasta este tema hemos visto como analizar la web social utilizando SNA. El objetivo de este tema es completar los análisis realizados utilizando métodos de minería de datos. De esta forma seremos capaces de encontrar modelos o patrones que expliquen la naturaleza de nuestros datos, los cuales son agrupaciones de información relevante debido a las interacciones de los usuarios de un sitio de la web social.

Este tema tiene un carácter principalmente práctico ya que lo importante es familiarizarse con las posibilidades que ofrece los métodos de minería de datos para representar y analizar lo que está sucediendo en un sitio de web social.

METODOLOGÍA

La metodología de esta asignatura se basará en dos paradigmas, por un lado, el *Aprendizaje colaborativo*, y por otro, *Aprender haciendo*. En esta asignatura se enseñará como analizar la web social, en donde hay relaciones entre distintos actores. Se da especial importancia, en el estudio y evaluación de la asignatura, a la realización de tareas prácticas, que podrán ser en equipo, analizando la web social.

Se utilizará de forma extensiva la plataforma virtual de la UNED, los cursos virtuales, como entorno virtual dinámico de interacción entre estudiantes y estos con el profesor.

Para favorecer la colaboración, se posibilitará la interacción de grupos pequeños de estudiantes mediante trabajos en equipo. Una vez empezado el curso, se abrirán subespacios en el curso virtual para mejorar la gestión de grupos pequeños de trabajo.

Los contenidos docentes se encuentran en la plataforma de e-learning de la UNED estructurados para la mejor comprensión y utilización por parte de los estudiantes.

Las actividades formativas de la asignatura son las que se listan a continuación:

•Estudios de contenidos (30 horas)

•Tutorías (5 horas)

•Actividades en la plataforma virtual (10 horas)

•Prácticas Informáticas (40 horas)

•Trabajos (15 horas)

•Total: 100 horas.

SISTEMA DE EVALUACIÓN

TIPO DE PRUEBA PRESENCIAL

Tipo de examen: Examen de desarrollo

Preguntas desarrollo

Duración del examen: 120 (minutos)

Material permitido en el examen

Ninguno

Criterios de evaluación

Criterios:

Que se haya respondido a todas las preguntas.

Que las respuestas demuestre con claridad los conocimientos del alumno dentro del área preguntada.

Que el alumno sea capaz de describir un proyecto de minería de datos en la web social.

% del examen sobre la nota final: 50

Nota del examen para aprobar sin PEC

Nota máxima que aporta el examen a la calificación final sin PEC

Nota mínima en el examen para sumar la PEC: 3

Comentarios y observaciones

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad: Si

Descripción

La única prueba presencial es el examen de desarrollo.

Criterios de evaluación

Ponderación de la prueba presencial y/o los trabajos en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC?: Si,PEC no presencial

Descripción

Las pruebas de evaluación continua (PEC) son obligatorias y tienen mucha importancia en la asignatura.

Las PEC se dividien en dos tareas, Tema 2 y 3, y una práctica en la que se podrá trabajar en equipo, Tema 4. La práctica tendrá un valor doble que las tareas.

Criterios de evaluación

Se evaluarán en las PEC:
Claridad en la exposición del trabajo realizado.
Alcanzar y sobrepasar las actividades encomendadas en los enunciados.
Relaccionar las conclusiones obtenidas con la teoría enseñada en la asignatura.
La original en la resolución del problema planteado.

Ponderación de la PEC en la nota final 40%
Fecha aproximada de entrega
Comentarios y observaciones

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? Si,no presencial
Descripción

La realización de otras actividades a lo largo del curso tendrá un valor en la evaluación final. Se tendrá en cuenta la participación activa y proactiva del estudiante en el espacio virtual, a través de los foros de la asignatura principalmente y la realización de tareas complementarias cuyos enunciados se detallarán en el curso virtual.

Criterios de evaluación

Ponderación en la nota final 10%
Fecha aproximada de entrega
Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

Nota final = (0.5 x [Nota del examen de desarrollo]) x (0,4 x [PEC]) x (0,1 x [Otras actividades])

BIBLIOGRAFÍA BÁSICA

ISBN(13):9780262012430
Título:INTRODUCTION TO MACHINE LEARNING (segunda)
Autor/es:Ethem Alpaydin ;
Editorial:Adaptive computation and machine learning

ISBN(13):9780761966524
Título:SOCIAL NETWORK ANALYSIS: THEORY AND APPLICATIONS
Autor/es:Varios Autores ;
Editorial:Creative Commons Attribution-Share Alike 3.0 Unported

En esta asignatura no se seguirá un libro para ir describiendo los contenidos del temario ya que la asignatura, por su carácter eminentemente práctico, abarca distintos contenidos teóricos. Sin embargo, los siguientes libros serán de gran utilidad:

- Social Network Analysis: Theory and Applications. Este libro ha sido escrito por muchos colaboradores siguiendo la licencia Creative Commons. En este libro se enseña en profundidad el análisis de redes sociales, ofreciendo toda la información necesaria para comenzar una análisis de este tipo. El libro se encuentra en formato pdf en la sección de Documentos del curso virtual.
- Social media mining: an introduction. De los autores: Zafarani, Reza, Mohammad Ali Abbasi, and Huan Liu. Publicado en 2014 por Cambridge University Press. De este libro es importante señalar como relaciona entre el análisis de redes sociales y técnicas de aprendizaje automático. El libro se encuentra en formato pdf en la sección de Documentos del curso virtual.
- Introduction to machine learning de Ethem Alpaydin editado por Adaptive computation and machine learning, 2010, segunda edición. Este trabajo se orienta a los estudiantes graduados con buena formación en matemáticas y algoritmos informáticos, así como las estadísticas. Alpaydin realiza un enfoque muy teórico, por lo tanto, los lectores deben estar preparados para trabajar a través de las fórmulas matemáticas y aplicar los algoritmos por su cuenta, que es la forma más eficaz para comprender el material. Los estudiantes que utilicen este libro tendrán una base sólida para explorar en profundidad más estudios como las redes neuronales, el aprendizaje supervisadas y no supervisadas, árboles de decisión, y otras áreas de este complejo campo. Referencias al final del libro ayudarán a los estudiantes a ampliar aún más su conocimiento de los temas presentados. Esta nueva edición (1ª ed., 2004) se revisó completamente e incluye varios nuevos capítulos y ejercicios. Es una contribución valiosa a una disciplina definida por investigadores importantes como Nils Nilsson, John McCarty, Marvin Minsky, Tom Mitchell, Stuart Russell y Peter Norwig. Este libro es especialmente interesante para el Tema 4, ya que en este tema se pondrán en práctica conocimientos que se estudian en la asignatura obligatoria de este master *Sistemas inteligentes y adquisición de conocimiento*, de la cual el libro es la bibliografía básica.

BIBLIOGRAFÍA COMPLEMENTARIA

Tema 1

1. Libro recopilatorio de artículos de investigación titulado Mining and Analyzing Social Networks, editado por I-Hsien Ting,Hui-JuWu, and Tien-Hwa Ho para Studies in Computational Intelligence,Volume 288.
2. Blog de Fernando Santamaría (http://fernandosantamaria.com/blog/).
3. Api para desarrolladores de Twitter (https://dev.twitter.com/overview/api).
4. Documentación para los desarrolladores de Facebook (https://developers.facebook.com/docs/).

Tema 2 y 3

1. Gephi, herramienta de análisis de redes sociales (https://gephi.org).
2. Introducción de Gephi al análisis de redes y visualización es muy recomendable (http://www.martingrandjean.ch/gephi-introduction/).

Tema 4

1. Weka, herramienta de análisis de datos utilizando aprendizaje automático (http://www.cs.waikato.ac.nz/ml/weka/).
2. R, herramienta de análisis de datos (https://www.r-project.org).

RECURSOS DE APOYO Y WEBGRAFÍA

Para alcanzar los objetivos propuestos, el curso se va a articular, como ya se ha comentado, a través de una plataforma especialmente diseñada para facilitar el trabajo individual y colaborativo en Internet (basada en comunidades virtuales), desarrollada por la Sección de Innovación del Centro de Innovación y Desarrollo Tecnológico de la UNED: aLF.
La plataforma de e-Learning aLF proporcionará el soporte requerido para gestionar los procesos de enseñanza y aprendizaje, compartir documentos y enlaces de interés, crear y participar en comunidades temáticas y grupos de trabajo específicos, realizar proyectos de diversa naturaleza, organizar el trabajo mediante agendas compartidas e individuales, acceder y publicar noticias de interés, etc. Esta plataforma permitirá realizar el seguimiento de las actividades del curso, así como estar al tanto de cualquier información o documentación de interés relacionada con el mismo.

ADENDA AL SISTEMA DE EVALUACIÓN CON MOTIVO DE LA PANDEMIA COVID 19

https://app.uned.es/evacaldos/asignatura/adendasig/31110107

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.