Analítica Web

Curso Académico: (2023 / 2024) **Fecha de revisión: 09-06-2021**

Departamento asignado a la asignatura: Departamento de Ingeniería Telemática

Coordinador/a: CUEVAS RUMIN, RUBEN

Tipo: Obligatoria Créditos ECTS : 6.0

Curso : 6 Cuatrimestre : 1

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Programación (Curso 1, Cuatrimestre 1)
Estructura de Datos y Algoritmos (Curso 1, Cuatrimestre 2)
Bases de Datos (Curso 2, Cuatrimestre 1)
Aplicaciones Web (Curso 3, Cuatrimestre 1)

OBJETIVOS

1. Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio.
2. Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio
3. Capacidad para diseñar soluciones basadas en aprendizaje automático en aplicaciones en dominios específicos como los sistemas de recomendación, el tratamiento del lenguaje natural, la Web o las redes sociales online.
4. Capacidad para desarrollar aplicaciones Web y móviles y utilizarlas para capturar datos con ellas.
5. Capacidad para utilizar herramientas de visualización de datos para poder comunicar los resultados de los análisis realizados sobre los mismos, adaptándolos a públicos diferentes tanto técnico como no técnico.
6. Conocimientos y habilidades adecuados para analizar y sintetizar problemas básicos relacionados con la ingeniería y la ciencia de datos, resolverlos y comunicarlos de forma eficiente.
7. Capacidad de resolver problemas con iniciativa, toma de decisiones, creatividad, y de comunicar y transmitir conocimientos, habilidades y destrezas, comprendiendo la responsabilidad ética, social y profesional de la actividad del tratamiento de datos. Capacidad de liderazgo, innovación y espíritu emprendedor.
8. Capacidad de comunicar los conocimientos oralmente y por escrito, ante un público tanto especializado como no especializado.
9. Haber adquirido conocimientos avanzados y demostrado una comprensión de los aspectos teóricos y prácticos y de la metodología de trabajo en el campo de la ciencias e ingeniería de datos con una profundidad que llegue hasta la vanguardia del conocimiento
10. Poder, mediante argumentos o procedimientos elaborados y sustentados por ellos mismos, aplicar sus conocimientos, la comprensión de estos y sus capacidades de resolución de problemas en ámbitos laborales complejos o profesionales y especializados que requieren el uso de ideas creativas e innovadoras
11. Tener la capacidad de recopilar e interpretar datos e informaciones sobre las que fundamentar sus conclusiones incluyendo, cuando sea preciso y pertinente, la reflexión sobre asuntos de índole social, científica o ética en el ámbito de su campo de estudio;

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

1. Colección de datos en el ecosistema Web:
   - 1.1 Arañas y scrappers web
   - 1.2 APIs
2. Analítica de datos en la Web
   - 2.1 Análisis de grafos: Métricas de centralidad e influencia
   - 2.2 Estructura de red:
     - 2.2.1 Tipos de redes (grafos bipartitos, redes mundo pequeño, redes de escala libre
     - 2.2.2 Clusterización, detección de comunidades, descomposición K-Core

3. Visualización de datos web
   3.1 Representación de información web.
   3.2 Herramientas de visualización.

4. Proyecto final de analítica web
   4.1 El Proyecto tiene que incluir los tres componentes presentados anteriormente (Colección de datos, Analítica de datos y Visualización de datos) en la Web.

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Esta asignatura estará basada en las siguientes actividades:

- CLASES TEORICO-PRACTICAS: Se introducirán los principales conceptos de la asignatura proporcionando ejemplos prácticos para facilitar el aprendizaje de los mismos. Se promoverá la participación de los estudiantes para discutir los conceptos y problemas introducidos en las clases de toería.

- CLASES DE LABORATORIO: En estas clases los estudiantes realizarán ejercicios prácticos de los conceptos introducidos en las clases de teoría. Los estudiantes deberán resolver problemas prácticos asociados a la analítica web.

-PROYECTO FINAL EN GRUPO: Se asignará a los estudiantes un proyecto que será realizado a lo largo del cuatrimestre en grupos de 2 o 3 personas. Los estudiantes deberán proponer su propio proyecto, en casos excepcionales los profesores podrán definir un listado de proyectos para que los estudiantes eligan de dicha lista en caso necesario. El profesor responsable deberá aprobar la propuesta de los estudiantes. El proyecto deberá incluir los siguientes elementos:

1- Una definición inicial de los objetivos del proyecto, la tecnología que se usará y los resultados esperados.
2- Utilizar alguna de las técnicas de colección de datas estudiadas para obtener información de algún servicio web o red social popular.
3- Análisis de los datos recogidos usando frameworks de análisis de datos actuales (por ejemplo R, Python, etc).
4- Visualización de los resultados.

Los estudiantes defenderán su proyecto en una exposición pública frente al resto de los estudiantes al final del cuatrimestre.

Se programarán varias clases en laboratorio que se usarán para supervisar la evolución del proyecto y para que los estudiantes progresen en su desarrollo.

TUTORIAS: Los alumnos tendrán acceso a tutorías individuales/colectivas todas las semanas para aclarar conceptos teóricos y/o prácticos y poder realizar un seguimiento más detallado de su proyecto.

SISTEMA DE EVALUACIÓN

- Practicas de laboratorio (50%): Las prácticas de laboratorio tendrán diferentes hitos a alcanzar. La consecución de los hitos irán aumentando la puntuación en la práctica hasta que en el caso de completar todos los hitos se obtendrá la máxima puntuación.
Se llevarán a cabo 4 prácticas de laboratorio. Cada una de ellas tiene un porcentaje en la nota final del 12,5%

- Proyecto Final (50%): La evaluación ser basará en la correcta implementación técnica del proyecto, el volumen de datos recogidos y procesados, la utilidad del proyecto y la eficiencia de la solución propuesta.

Para superar la asignatura hay que superar por separado las prácticas de laboratorio y el proyecto final. Es decir se deberá obtener un porcentaje mínimo de 20% sobre el total de la nota en las prácticas de laboratorio y un porcentaje mínimo de 20% sobre el total de la nota en el proyecto final.

Peso porcentual del Examen Final:	0
Peso porcentual del resto de la evaluación:	100

BIBLIOGRAFÍA BÁSICA

- Christopher Olston, Marc Najork. Web Crawling, Now Publishers Inc, , 2010

- Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman Mining of massive datasets, Cambridge University

Press., 2014

- Stanley Wasserman, Katherine Faust Social Network Analysis: Methods and Applications, Cambridge University Press., 1994

RECURSOS ELECTRÓNICOS BÁSICOS

- Jure Leskovec . (Social and Information) Network Analysis: http://snap.stanford.edu/na09/resources.html