

A10. Big Data: Herramientas para el procesamiento de datos masivos

**MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN
INTELIGENCIA ARTIFICIAL**

UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.



DATOS GENERALES

Breve descripción

En asignatura se aprenderán las principales herramientas disponibles para trabajar con bases de datos masivas. Veremos las diferentes técnicas y algoritmos de analítica de datos, desde el preprocesado a la clasificación, tratamiento de datos en streaming, etc.

Título asignatura

A10. Big Data: Herramientas para el procesamiento de datos masivos

Código asignatura

102126

Curso académico

2017-18

Planes donde se imparte

[MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL](#)

Créditos ECTS

4,5

Carácter de la asignatura

OPTATIVA

Duración

Anual

Idioma

Castellano

CONTENIDOS

Contenidos

Los avances tecnológicos de los últimos años han modificado nuestros hábitos y estilos de vida de una manera difícil de imaginar hace pocos años. El desarrollo de internet y su popularidad a nivel mundial han ayudado a eliminar fronteras y han creado multitud de servicios donde los datos transmitidos son un eje central de su funcionamiento. No obstante, estos datos no sólo se encuentran ligados a internet o a las redes sociales, sino que son parte fundamental de numerosas aplicaciones, tales como las colecciones de datos que nos proporcionan los instrumentos científicos, las redes de sensores, los dispositivos móviles, las transacciones comerciales, la genómica y la biomedicina, o los sistemas de información de la empresa.

Esta gran cantidad de datos disponible en la actualidad y las tecnologías necesarias para su procesamiento conforma lo que conocemos hoy día como "big data". Esta materia se centrará en el procesamiento de datos masivos, tanto en los principios formales como en las herramientas específicas para tratar estos volúmenes de datos.

- Big data.
- Procesamiento de datos masivos.
- Deep learning.
- Herramientas para el tratamiento de grandes volúmenes de datos: Hadoop, Spark, Mahout, MLlib.

Unidades

1. Módulo 1: Fundamentos de Big Data. Algunas aplicaciones
2. Módulo 2: Modelo de programación MapReduce
3. Módulo 3: Hadoop. Un caso de estudio
4. Módulo 4: Analítica para Big data. Generalidades y herramientas
5. Módulo 5: Algoritmos de Preprocesamiento
6. Módulo 6: Algoritmos de clasificación
7. Módulo 7: Algoritmos de Asociación
8. Módulo 8: Data streaming

9. Módulo 8: Herramientas Big data

COMPETENCIAS

Generales

CG1 - Entender los conceptos, los métodos y las aplicaciones de la inteligencia artificial.

CG2 - Evaluar nuevas herramientas computacionales y de gestión del conocimiento en el ámbito de la Inteligencia Artificial.

CG3 - Gestionar de manera inteligente los datos, la información y su representación.

Específicas

CE2 - Aplicar las técnicas de aprendizaje automático utilizando la metodología de validación y presentación de resultados más apropiada en cada caso.

CE5 - Analizar las fuentes documentales propias del ámbito de la investigación en Inteligencia Artificial para poder determinar cuáles de ellas son relevantes en la resolución de problemas concretos.

PLAN DE APRENDIZAJE

Actividades formativas

A1 - **Sesiones presenciales virtuales**: visionado inicial del material audiovisual (vídeos introductorios, presentaciones, animaciones) que se elabore en cada una de las materias y que servirán presentación de cada uno de los temas a los estudiantes (12 horas - 100% presencialidad).

A2 - **Trabajos individuales**: realización de ejercicios, resolución de problemas, realización de prácticas y/o trabajos/proyectos individuales (17 horas - 0% presencialidad).

A3 - **Trabajo autónomo**: estudio del material básico, lecturas complementarias y otros contenidos y estudio (72 horas - 0% presencialidad).

A4 - **Foros y chats**: lanzamiento de cuestiones y temas para la discusión general (5,5 horas - 0% presencialidad).

A5 - **Tutorías**: consultas y resolución de dudas, aclaraciones, etc (6 horas - 100% presencialidad).

SISTEMA DE EVALUACIÓN

Descripción del sistema de evaluación

E1 - Valoración de los cuestionarios de evaluación: los estudiantes realizarán por cada unidad didáctica un cuestionario de evaluación que será objeto de puntuación en la nota final (ponderación mínima 20% y máxima 40%).

E2 - Valoración de la participación en foros y chats: se valorará el nivel de participación/debate de los estudiantes que contará para la nota final (ponderación mínima 10% y máxima 20%).

E3 - Valoración de los trabajos individuales: se valorarán los problemas, proyectos, trabajos realizados y entregados a través de la plataforma, y apoyado en los casos que sea necesario (sobre todo cuando se trate de desarrollo de código) por plataformas de gestión de código como GitHub. También se incluirá el video que el alumno deberá enviar al profesor para cada asignatura (ponderación mínima 40% y máxima 70%).

Calendario de exámenes

Para la **convocatoria ordinaria**, habrá 3 fechas de entrega de trabajos final de curso. Los alumnos podrán entregar sus trabajos en cualquier momento, pero sólo en estas fechas se recogerán y evaluarán los que se hayan entregado.

Las fechas serán:

- 20/12/17
- 27/2/18
- 31/5/18

Habrà una **convocatoria extraordinaria** en todas las asignaturas. Para su evaluación, la entrega de trabajos se cerrará el día:

- 01/07/18

PROFESORADO

Profesor responsable

Alonso Betanzos, María Amparo

*Catedrática de Ciencias de la Computación e Inteligencia Artificial
Universidad de A Coruña*

Profesorado

Martínez Rego, David

*Doctor en Informática
Universidad de A Coruña*

Eiras Franco, Carlos

*Profesor Ayudante Doctor
Universidad de A Coruña*

Bolón Canedo, Verónica

*Profesora Titular de Universidad
Universidad de A Coruña*

Seijo Pardo, Borja

*Teaching Assistant
Investigador predoctoral en Ciencias de la Computación
Universidad de A Coruña*

Morán Fernández, Laura

*Teaching Assistant
Personal Investigador Predoctoral en Formación
Universidad de A Coruña*

HORARIO

Horario

Todas las asignaturas estarán en la plataforma a disposición de los estudiantes desde octubre hasta julio.

BIBLIOGRAFÍA Y ENLACES RELACIONADOS

Bibliografía

Sean T. Allen, Matthew Jankowski, and Peter Pathirana. *Storm Applied*. Manning 2015

Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia. *Learning Spark*. O'Reilly 2015

Sameer B. Wadkar, Hari Rajaram. *Flink in Action*. Manning 2017

Paul Butcher. *Seven concurrency models in seven weeks. The Pragmatic Programmer* 2014

Mahmoud Parsian. *Data Algorithms: Recipes for Scaling Up with Hadoop and Spark*. O'Reilly 2015

Tom White. *Hadoop: The Definitive Guide*, 4th Edition. O'Reilly 2015

Thilina Gunarathne. *Hadoop MapReduce v2 Cookbook*, 2nd Edition. Packt Publishing, 2015

Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia. *Learning Spark Lightning-Fast Big Data Analysis*. O'Reilly Media, 2015

Venkat Ankam. *Big Data Analytics*. Packt Publishing, 2016.

Vladimir Bacvanski. *Introduction to Big Data An Overview of Fundamental Big Data Concepts, Tools, Techniques and Practices*. O'Reilly Media, 2015.

Verónica Bolón-Canedo, Noelia Sánchez-Marroño, and Amparo Alonso-Betanzos. *Feature selection for high-dimensional data*. Springer, 2015.

Salvador García, Julián Luengo, and Francisco Herrera. *Data preprocessing in data mining*. New York: Springer, 2015.

Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti Zadeh, Eds. *Feature Extraction: Foundations and Applications*. Springer, 2006.