



Curso Académico 2017-18

HERRAMIENTAS INFORMÁTICAS PARA BIG-DATA

Ficha Docente

ASIGNATURA

Nombre de asignatura (Código GeA): HERRAMIENTAS INFORMÁTICAS PARA BIG-DATA (608857)

Créditos: 3

Créditos presenciales: 3

Créditos no presenciales:

Semestre: 2

PLAN/ES DONDE SE IMPARTE

Titulación: MÁSTER EN TRATAMIENTO ESTADÍSTICO COMPUTACIONAL DE LA INFORMACIÓN
Plan: MÁSTER UNIVERSITARIO EN TRATAMIENTO ESTADÍSTICO COMPUTACIONAL DE LA INFORMA
Curso: 1 **Ciclo:** 2
Carácter: OPTATIVA
Duración/es: Por determinar (no genera actas)
Idioma/s en que se imparte:
Módulo/Materia: /

PROFESOR COORDINADOR

Nombre	Departamento	Centro	Correo electrónico	Teléfono
--------	--------------	--------	--------------------	----------

PROFESORADO

Nombre	Departamento	Centro	Correo electrónico	Teléfono
GREGORIO RODRIGUEZ, CARLOS	Sistemas Informáticos y Computación	Facultad de Estudios Estadísticos	cgregor@ucm.es	

SINOPSIS

BREVE DESCRIPTOR:

La asignatura plantea una introducción a los conceptos y técnicas utilizadas en el Análisis de Datos en ámbitos de Big Data.

Se plantea una visión panorámica de las tecnologías utilizadas en las diversas fases: obtención de la información, almacenamiento, procesamiento, explotación y visualización.

REQUISITOS:

Haber cursado, al menos, un primer curso de programación.

OBJETIVOS:

- Entender la distintas fases del Análisis de datos
- Conocer herramientas informáticas para aplicar en cada una de las fases
- Comprender el marco conceptual del BigData
- Entender y manejar básicamente un sistema HDFS Hadoop
- Diseñar soluciones paralelizables utilizando el esquema Map Reduce
- Escribir programas en Spark

COMPETENCIAS:

Generales:

CG1 – Aprender a aplicar los conocimientos adquiridos y a explotar su potencial para la resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) en el tratamiento estadístico computacional de la información.

CG2 – Elaborar adecuadamente y con originalidad argumentos motivados y proyectos de trabajo, redactar planes, así como formular hipótesis y conjeturas razonables en su área de especialización.

CG3 – Integrar los conocimientos adecuados y enfrentarse a la complejidad de emitir juicios en función de criterios, de normas externas o de reflexiones personales justificadas.

CG5 – Comprender y utilizar el lenguaje y las herramientas matemáticas para modelizar y resolver problemas complejos, reconociendo y valorando las situaciones y problemas susceptibles de ser tratados matemáticamente.



Curso Académico 2017-18

HERRAMIENTAS INFORMÁTICAS PARA BIG-DATA

Ficha Docente

CG6 – Conocer los modelos, métodos y técnicas relevantes en distintas áreas de aplicación de la Estadística matemática participando en la creación de nuevas tecnologías que contribuyan al desarrollo de la Sociedad de la Información.

Transversales:

CT1 – Saber aplicar sus conocimientos a su trabajo o vocación de una forma profesional y poseer las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y en la resolución de problemas y estudio de casos. Esto implica, más concretamente: Integrar creativamente conocimientos y aplicarlos a la resolución de problemas complejos, perseguir objetivos de calidad en el desarrollo de su actividad profesional, adquirir capacidad para la toma de decisiones y de dirección de recursos humanos, ser capaz de mostrar creatividad, iniciativa y espíritu emprendedor para afrontar los retos de su actividad, valorar la importancia de los métodos estadístico-computacionales en el contexto industrial, económico, administrativo, medio ambiental y social.

CT2 – Tener la capacidad de reunir e interpretar datos relevantes para emitir juicios que incluyan una reflexión sobre temas relevantes de índole científica, tecnológica y empresarial. Demostrar razonamiento crítico y gestionar información científica y técnica de calidad, bibliografía, bases de datos especializadas y recursos accesibles a través de Internet.

Específicas:

CE1 – Adquisición de una formación sólida y rigurosa en temas avanzados de Estadística Matemática y Tecnologías de la Decisión aplicadas al tratamiento de la Información.

CE2 – Capacidad para planificar la resolución de un problema en función de las herramientas de que se disponga y, en su caso, de las restricciones de tiempo y recursos.

CE3 – Capacidad para utilizar aplicaciones informáticas estadísticas, de cálculo numérico y simbólico, visualización gráfica, optimización u otras para resolver problemas con un elevado grado de complejidad.

CE4 – Desarrollar habilidades de aprendizaje en Estadística Computacional y Matemáticas, así como en sus respectivas aplicaciones, que permitan al alumno continuar estudiando y profundizando en la materia de modo autónomo, así como el desarrollo profesional con un alto grado de independencia.

CE5 – Resolver problemas y casos reales planteados en el tratamiento estadístico computacional de la información generada en los ámbitos de la ciencia, la tecnología y la sociedad mediante habilidades de modelización matemática, estimación y computación.

CE6 – Desarrollar programas que resuelvan problemas matemáticos utilizando para cada caso el entorno computacional adecuado.

CE7 – Capacidad de utilización de herramientas de búsqueda de recursos bibliográficos así como manejo, gestión y análisis de grandes bases de datos.

Otras:

CONTENIDOS TEMÁTICOS:

p { margin-bottom: 0.25cm; direction: ltr; line-height: 120%; text-align: left; }

- Introducción a Big Data.
- Fuentes de datos de información en la web. Gestión de formatos: CSV, XML y JSON.
- Introducción al análisis de datos utilizando Python: Pandas, Numpy y Matplotlib.
- Almacenamiento de datos en ambientes Big Data.
 - Bases de Datos SQL, NoSQL y New SQL.
 - Sistemas de Ficheros distribuidos Hadoop HDFS
- Técnicas de procesamiento paralelo:
 - Hadoop Map Reduce
 - Spark.
- Herramientas de Inteligencia de Negocio: Pentaho, Tableau

ACTIVIDADES DOCENTES:

Clases teóricas:

En las que se expondrán los conceptos y técnicas esenciales de cada tema.

Clases prácticas:

En las que se resolverán ejercicios y prácticas relacionadas con los conceptos y técnicas de cada tema.

Trabajos de campo:



Curso Académico 2017-18

HERRAMIENTAS INFORMÁTICAS PARA BIG-DATA

Ficha Docente

Prácticas clínicas:

Laboratorios:

En las que se utilizarán herramientas actualmente utilizadas tanto en el ámbito académico como profesional.

Exposiciones:

Presentaciones:

Otras actividades:

TOTAL:

EVALUACIÓN

La evaluación de la asignatura es continua con entregas de trabajos a medida que avanza el curso.

BIBLIOGRAFÍA BÁSICA:

L. Massaron; A. Boschetti, Python Data Science Essentials, Packt. 2015

J. VanderPlas, Python Data Science Handbook, O'Reilly. 2016

Documentación Python: <https://docs.python.org>

Documentación Hadoop: <http://hadoop.apache.org/docs/current/>

Documentación MrJob: <https://pythonhosted.org/mrjob/>

Documentación Spark: <http://spark.apache.org/docs/latest/>

Karau H., Konwinski A., Wendell P., and Zaharia M. Learning Spark. O'Reilly. 2015

OTRA INFORMACIÓN RELEVANTE: