

Guía docente de Asignatura–Máster en Minería de Datos e Inteligencia de Negocios

Datos generales de la asignatura

Asignatura: Gestión y Explotación de Almacenes de Datos - 606542

Curso académico: 2019-20

Carácter: Obligatoria

Semestre: 2

Créditos ECTS

Presenciales: 4,5

No presenciales: 1,5

Total 6,0

Actividades docentes

Clases teóricas: 50%

Clases prácticas: 50%

Total: 100%

Departamentos responsables: Departamento de Sistemas Informáticos y Computación

Profesores: Ricardo Rodríguez Cielos / Cesar Tejedor Moreno

Datos específicos de la asignatura

Breve descriptor: Estudio de técnicas para la obtención de conocimiento a partir de bases de datos relacionales y aplicación de estas técnicas a diferentes campos científicos y empresariales.

Requisitos: Haber cursado una asignatura de bases de datos que incluya el estudio del modelo relacional y el uso práctico del lenguaje SQL.

Competencias

Generales: En este curso se estudian los almacenes de datos desde el punto de vista de su aplicación en técnicas básicas tanto de análisis como de minería de datos (v.g. asociación, predicción o clustering), junto con sus características, limitaciones y posibles ámbitos de implantación (por ejemplo, el ámbito científico, el empresarial o el financiero). Para ello, se usan herramientas informáticas tales como Weka, Knime o similares.

Contenidos

- Repaso de bases de datos relacionales, lenguaje SQL.
- Introducción al lenguaje Python: acceso a bases de datos. Recopilación de los datos en la Web/webcrawling). Acceso a datos de redes sociales.
- Esquema de programación Map Reduce
- Big Data :Hadoop, HDFS y Spark
- Ajustes para la ejecución distribuida de tareas

Evaluación

La evaluación consistirá en la resolución de problemas propuestos por el profesor a lo largo del curso. Este método sirve tanto para convocatoria de junio como de septiembre.

Bibliografía

- Parker, James R, Python: an introduction to programming. Dulles (Virginia, EEUU): Mercury Learning and Information, cop.2017
- Russell; Matthew A. Mining the social web: [Data mining Facebook, Twitter, LinkedIn, Google+, GitHub, and more]/Matthew A. Russell. Sebastopol (California): O'Really media, 2013
- Documentación Hadoop: <http://hadoop.apache.org/docs/current/>
- Documentación MrJob: <http://pythonhosted.org/mrjob/>
- Karau H, Konwinski A., Wendell P., and Zaharia M. Learning Spark. O'Really. 2015
- Documentación Spark: <http://spark.apache.org/docs/latest/>