



Proyecto docente de la asignatura

Asignatura	TÉCNICAS ESCALABLES DE ANÁLISIS DE DATOS		
Materia	SISTEMAS INTELIGENTES Y BASADOS EN CONOCIMIENTO		
Módulo			
Titulación	MÁSTER EN INGENIERÍA INFORMÁTICA		
Plan	510	Código	53198
Periodo de impartición	S2	Tipo/Carácter	OB
Nivel/Ciclo	MÁSTER	Curso	1
Créditos ECTS	6		
Lengua en que se imparte	ESPAÑOL		
Profesor/es responsable/s	CARLOS J. ALONSO GONZÁLEZ, BELARMINO PULIDO JUNQUERA, PEDRO C. ÁLVAREZ ESTEBAN		
Departamento	INFORMÁTICA (ATC, CCIA, LSI), ESTADÍSTICA E INVESTIGACIÓN OPERATIVA		
Datos de contacto (E-mail, teléfono...)	calonso@infor.uva.es 983 185602; belar@infor.uva.es 983 185606; pedroc@eio.uva.es 983 423930		



4. Contenidos

Bloque 1: Técnicas Escalables de Análisis de Datos.

Carga de trabajo en créditos ECTS:

6

1. Introducción a Apache Spark.
 - a. Spark y sus componentes.
 - b. Introducción a Scala.
 - c. RDDs y DataFrames.
2. Conceptos generales sobre Aprendizaje Automático y Grandes Volúmenes de Datos.
 - a. Arquitectura de un sistema de aprendizaje automático.
 - b. Acceso, procesamiento y filtrado de datos.
3. Métodos de aprendizaje sobre ML/ MLLib / Spark.
 - a. Aprendizaje supervisado: Clasificadores y Modelos de Regresión.
 - b. Aprendizaje no supervisado: Técnicas de agrupamiento (*clustering*).
 - c. Recomendadores.

Se mantienen los contenidos inicialmente previstos porque se considera que son los mínimos para realizar los proyectos con los que se evalúa la asignatura. A cambio, se reduce la intensidad de los proyectos con los que se evalúa la asignatura.

Bibliografía básica

- Nick Pentreath. Machine Learning with Spark. Packt Publishing. 2015. ISBN: 9781783288519. <http://www.packtpub.com/>
- Petar Zečević y Marko Bonaći. Spark in Action. Manning Publications. 2016. ISBN: 9781617292606. <https://www.manning.com/books/>
- Ian H. Witten, Eibe Frank y Mark A. Hall. Data Mining: practical machine learning tools and techniques (third Edition). Morgan Kaufmann, 2011.

Bibliografía complementaria

- Apache Organization. Apache Spark. <http://spark.apache.org/>
- Apache Organization. Apache MLLib. <http://spark.apache.org/mllib/>
- Kaggle. Kaggle in class. <https://inclass.kaggle.com/>
- Rishi Yadav. Spark Cookbook. Packt Publishing 2015.
- Jure Leskovek, Anand Rajaraman, Jeffrey D. Ullman. Mining of Massive Datasets. Second edition. Cambridge University Press, 2014.



5. Métodos docentes y principios metodológicos

Clase magistral participativa online para discutir los contenidos básicos de la asignatura.

Laboratorios para la experimentación con las ideas básicas del bloque temático: los estudiantes disponen de Máquinas Virtuales que pueden usar de forma remota y las dudas se resuelven mediante el uso de mensajes en Moodle o tutorías online.

Realización de proyectos.

Recursos necesarios

Se usará LifeSize para las clases magistrales y las tutorías asociadas a los laboratorios.

Notas de la asignatura.

Guiones de cuestiones y problemas.

Curso Moodle de soporte a la asignatura: tanto para publicar contenidos como chats para dudas y mensajes.

Software de libre disposición para el desarrollo de análisis de datos escalable como Spark.

6. Tabla de dedicación del estudiante a la asignatura

ACTIVIDADES PRESENCIALES /	HORAS	ACTIVIDADES NO PRESENCIALES / NO PRESENCIALES ON LINE	HORAS
Clases teórico-prácticas (T/M)	5	Estudio y trabajo autónomo individual	30
		Clases teórico-prácticas online (T/M)	10
Laboratorios (L)	10	Estudio y trabajo autónomo individual	24
		Laboratorios online (L)	20
Seminarios (S)	5	Estudio y trabajo grupal dirigido	28
		Seminarios online (S)	10
Evaluación*			8
Total presencial	30	Total no presencial	120

* Evaluación: Se incluyen en las actividades de Laboratorio y Seminarios.



7. Sistemas y características de la evaluación

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Entregables	10%	Se realizarán dos o tres entregables: sobre Scala/Spark, sobre clasificadores y/o sobre regresión. Todos los entregables se realizaban como una tarea en el Moodle de la titulación, por lo tanto, no ha habido modificación.
Proyectos y Mini-proyectos	85%	Se realizarán dos mini-proyectos al finalizar las partes de clasificación y regresión. Su peso conjunto será del 40% de la nota final. Finalmente se realizará un proyecto que supondrá el 45% de la nota final que integrará varias de las técnicas vistas. Todas las entregas de proyectos se realizaban como tareas en el Moodle de la titulación, por lo tanto, no ha habido modificación.
Participación en clases, cuestionarios, seminarios prácticas y tutorías.	5%	La participación en clases, seminarios, prácticas y tutorías se evalúa a partir de la participación en las clases y laboratorios, tanto presenciales como online.

CRITERIOS DE CALIFICACIÓN

Se mantienen los criterios de evaluación dado que ya se usaba evaluación continua y todas las entregas se realizaban ya online.

- **Convocatoria ordinaria:**
 - Se realizará evaluación continua con las actividades y pesos indicados en la tabla anterior. Se exigirá obtener un mínimo de cuatro en las distintas pruebas para poder hacer la nota media.
- **Convocatoria extraordinaria:**
 - La calificación del 100% de la nota de la convocatoria extraordinaria se obtendrá mediante la realización de un proyecto. No obstante, aquellos estudiantes que quieran conservar las calificaciones obtenidas en las partes de Entregables y/o Mini-Proyectos de la convocatoria ordinaria (máximo de 50% de la nota en evaluación continua) podrán solicitarlo con antelación y en ese caso **se complementará la nota hasta llegar al 100% con la nota del proyecto.**

8. Consideraciones finales