



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 | INFORMATICA (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 | | |



1. Situación / Sentido de la Asignatura

1.1 Contextualización

La asignatura "Técnicas Escalables de Análisis de Datos" introduce los elementos necesarios para aplicar técnicas de Aprendizaje Automático a grandes volúmenes de datos como lo son los procedentes de aplicaciones web o móviles, la Internet de las Cosas y las redes de sensores, así como procedentes de servicios financieros, sanidad u otros campos científicos.

El conjunto de datos que se puede usar en estos campos es enorme y el conjunto de técnicas de aprendizaje a aplicar muy variado. Estos datos puede ser propiedad de una organización o pueden proceder de múltiples fuentes, pero en todos los casos su volumen puede ser tan grande que no se puedan procesar en un único ordenador, por lo cual será necesario recurrir posiblemente a un almacenamiento distribuido, a un procesamiento distribuido o a ambos.

Además, la gran cantidad de datos a procesar hará necesario analizar con cuidado el tipo de técnicas o algoritmos aplicables, ya que los requisitos de memoria pueden hacer inviables la utilización de técnicas o aplicaciones más convencionales.

1.2 Relación con otras materias

Existe relación con las asignaturas de la especialidad de Big Data como son la "Tecnología para el Big Data", "Almacenamiento Escalable" y "Big Data: Inteligencia de Negocios", donde se tratarán problemas asociados a la gestión y almacenamiento de grandes cantidades de datos en entornos distribuidos y su posterior uso en los procesos de negocio para la extracción de información con técnicas similares a la Minería de Datos o el Descubrimiento de Conocimiento en Bases de Datos.

El aprendizaje de modelos a partir de grandes cantidades de datos también está relacionado con el aprendizaje de modelos gráficos probabilísticos como son las Rede Bayesianas o las Redes de Markov que se estudian en la asignatura obligatoria del máster "Métodos avanzados de razonamiento y representación del conocimiento".

1.3 Prerrequisitos

Se recomienda que el alumno haya cursado estudios de grado con un contenido medio de competencias en Inteligencia Artificial y en Matemática Discreta. En relación con los Grados de Informática hasta ahora vigentes en los planes de estudio de la UVa, se recomienda que el alumno haya cursado la asignatura de "Técnicas de Aprendizaje Automático".



2. Competencias

2.1 Generales

| Código | Descripción |
|--------|---|
| CG1 | Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la ingeniería informática. |
| CG3 | Capacidad para dirigir, planificar y supervisar equipos multidisciplinares. |
| CG4 | Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática. |
| CG7 | Capacidad para la puesta en marcha, dirección y gestión de procesos de fabricación de equipos informáticos, con garantía de la seguridad para las personas y bienes, la calidad final de los productos y su homologación. |
| CG8 | Capacidad para la aplicación de los conocimientos adquiridos y de resolver problemas en entornos nuevos o poco conocidos dentro de contextos más amplios y multidisciplinares, siendo capaces de integrar estos conocimientos. |
| CG9 | Capacidad para comprender y aplicar la responsabilidad ética, la legislación y la deontología profesional de la actividad de la profesión de Ingeniero en Informática. |

2.2 Específicas

| COMPETENCIAS ESPECÍFICAS | |
|--------------------------|---|
| Código | Descripción |
| CET5 | Capacidad para analizar las necesidades de información que se plantean en un entorno y llevar a cabo en todas sus etapas el proceso de construcción de un sistema de información. |
| CET9 | Capacidad para aplicar métodos matemáticos, estadísticos y de inteligencia artificial para modelar, diseñar y desarrollar aplicaciones, servicios, sistemas inteligentes y sistemas basados en el conocimiento. |

3. Objetivos

| Código | Descripción |
|---------|---|
| SI-BC 5 | Conocer los métodos básicos de aprendizaje automático y Minería de Datos y entender los problemas de escalabilidad en entornos de grandes almacenes de datos, como es el caso de los entornos Big Data. |
| SI-BC 6 | Conocer cómo se pueden implementar estas técnicas en frameworks específicos, así como sus limitaciones. |
| SI-BC 7 | Conocer y ser capaz de aplicar técnicas de análisis de datos mediante clasificación (tanto supervisada como no supervisada), regresión, agrupamiento (clustering) o asociación. |
| SI-BC 8 | Ser capaz de aplicar técnicas avanzadas de análisis de datos como regresión, recomendadores, clustering o clasificación. |
| SI-BC 9 | Ser capaz de implementar estas técnicas en distintos ámbitos de aplicación, utilizando las tecnologías adecuadas |



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.

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 para discutir los contenidos básicos de la asignatura.

Laboratorios para la experimentación con las ideas básicas del bloque temático.

Realización de proyectos.

Recursos necesarios

Notas de la asignatura.

Guiones de cuestiones y problemas.

Curso Moodle de soporte a la asignatura.

Software gratuito 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 | HORAS |
|---|-----------|---------------------------------------|-----------|
| Clases teórico-prácticas (T/M) | 15 | Estudio y trabajo autónomo individual | 30 |
| Clases prácticas de aula (A) | | | |
| Laboratorios (L) | 30 | Estudio y trabajo autónomo individual | 24 |
| Prácticas externas, clínicas o de campo | | | |
| Seminarios (S) | 15 | Estudio y trabajo grupal dirigido | 28 |
| Tutorías grupales (TG) | | | |
| Evaluación* | | | 8 |
| Total presencial | 60 | Total no presencial | 90 |

* 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. |
| 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. |
| Participación en clases, cuestionarios, seminarios prácticas y tutorías. | 5% | La participación en clases, seminarios, prácticas y tutorías se evalúa a partir de las entregas opcionales y la participación en clase. |

CRITERIOS DE CALIFICACIÓN

- **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 sólo tendrían que realizar la parte proporcional del proyecto final (50%, 65% u 85% según la parte de la nota que se quiera conservar).

Recuerde que, aunque en ningún caso la asistencia a clase es evaluable, los profesores responsables pueden excluir de alguna actividad formativa evaluable a aquellos alumnos que no participen en las actividades presenciales, que incluyen las tutorías activas, los seminarios y las prácticas de laboratorio, especialmente, aunque no limitado a, en aquellas actividades de carácter grupal.

8. Consideraciones finales