

# Minería de Datos

---

***José Hernández Orallo***

**jorallo@dsic.upv.es**

**Máster y Cursos de Postgrado del DSIC  
Universitat Politècnica de València**

## Objetivos Generales

---

El objetivo fundamental de este curso es:

saber aplicar las técnicas de minería de datos apropiadas para problemas concretos de extracción de conocimiento útil para el análisis o la toma de decisiones.

2

## Objetivos Específicos (1/2)

---

- Reconocer la problemática del análisis de grandes volúmenes de datos y de los beneficios de su uso.
- Conocer las fases del Descubrimiento de Conocimiento de Bases de Datos y la importancia de las mismas en el éxito del proceso.
- Conocer básicamente las distintas técnicas de aprendizaje automático y estadísticas utilizadas en Minería de Datos (Data Mining), su potencial, su coste computacional y sus limitaciones de representación y de inteligibilidad.

3

## Objetivos Específicos (2/2)

---

- Elegir, para un problema concreto, qué técnicas de minería de datos son más apropiadas y utilizarlas, mediante una herramienta o paquete de minería de datos.
- Evaluar la calidad de un modelo, utilizando técnicas sencillas de evaluación (validación cruzada).
- Conocer la problemática especial de la Minería Web (documentos textuales e hipertextuales, XML) y las técnicas más usuales.

4

## Temario

---

1. Introducción a la Minería de Datos (DM)
  - 1.1. Motivación
  - 1.2. Problemas tipo y aplicaciones
  - 1.3. Relación de DM con otras disciplinas
2. El proceso de KDD
  - 2.1. Las Fases del KDD
  - 2.2. Tipología de Técnicas de Minería de Datos
  - 2.3. Sistemas Comerciales y Herramientas de Minería de datos
  - 2.4. Preparación y Visualización de datos
3. Técnicas de Minería de Datos
  - 3.1. El Problema de la Extracción Automática de Conocimiento.
  - 3.2. Evaluación de Hipótesis
  - 3.3. Técnicas no supervisadas y descriptivas.
  - 3.4. Técnicas supervisadas y predictivas.
4. Web Mining
  - 4.1. Los Problemas de la Información No Estructurada.
  - 4.2. Extracción de Conocimiento a partir de Documentos HTML y texto.
  - 4.3. Extracción de Información semi-estructurada (XML).
5. Otros Aspectos

5

## Material y Organización

---

### Material:

- Transparencias y boletines.
- Software en el laboratorio:  
SPSS Clementine 9.0
- Otra Documentación:

<http://www.dsic.upv.es/~jorallo/master>

### Sesiones de 4 horas:

- 1,5 exposición
- 2,5 prácticas con Clementine

6

## Bibliografía (1/2)

---

- Berry M.J.A.; Linoff, G.S. "Mastering Data Mining" Wiley 2000.
- Berthold, M.; Hand, D.J. (ed) "Intelligent Data Analysis. An Introduction" 2nd Edition, Springer 2002.
- Dunham, M.H. "Data Mining. Introductory and Advanced Topics" Prentice Hall, 2003.
- Dzeroski, S.; Lavrac, N. "Relational Data Mining" Springer 2001.
- Etzioni, O. "The World- Wide Web. Quagmire or Gold Mine" Communications of the ACM, November 1996, Vol. 39, nº 11, 1996.
- Fayyad, U.M.; Piatetskiy-Shapiro, G.; Smith, P.; Ramasamy, U. "Advances in Knowledge Discovery and Data Mining", AAAI Press / MIT Press, 1996.
- Fayyad, U.M.; Grinstein, G.; Wierse, A. "Information Visualization in Data Mining and Knowledge Discovery" Morgan Kaufmann, Harcourt Intl., 2001.
- Han, J.; Kamber, M. "Data Mining: Concepts and Techniques" Morgan Kaufmann, 2001.

7

## Bibliografía (2/2)

---

- Hand, D.J.; Mannila, H. and Smyth, P. "Principles of Data Mining", The MIT Press, 2000.
- Hernández, J.; Ramírez, M.J.; Ferri, C. "Introducción a la Minería de Datos" Pearson Prentice Hall, 2004.
- Kosala, R.; Blockeel, H. "Web Mining Research: A Survey" ACM SIGKDD Explorations, Newsletter of the ACM SIG on Knowledge Discovery and Data Mining, June 2000, Vol. 2, nº1, pp. 1-15.
- Mena, Jesus "Data Mining Your Website", Digital Press, July 1999.
- Mitchell, T.M. "Machine Learning" McGraw-Hill 1997.
- Pyle, D. "Data Preparation for Data Mining" Morgan Kaufmann, Harcourt Intl., 1999.
- Thuraisingham, B. "Data Mining. Technologies, Techniques, Tools, and Trends", CRC Press, 1999.
- Witten, I.H.; Frank, E. "Tools for Data Mining", Morgan Kaufmann, 2nd Edition, 2004.
- Wong, P. C. "Visual Data Mining", Special Issue of *IEEE Computer Graphics and Applications*, Sep/Oct 1999, pp. 20- 46.

8