

## ALMACENES DE DATOS OLAP Y MINERÍA DE DATOS

|                                    |                             |
|------------------------------------|-----------------------------|
| <b>Código: MIC 108</b>             | <b>Obligatoria</b>          |
| <b>Horas Módulo: 40 (cuarenta)</b> | <b>Créditos: 2.6</b>        |
| <b>Área: Especializada (S-D)</b>   | <b>Correlativa:</b>         |
| <b>Horas teóricas: 50%</b>         | <b>Horas prácticas: 50%</b> |

### 1. FUNDAMENTACIÓN

En este curso se presentará el problema del análisis de sistemas de información para la toma de decisiones y las herramientas genéricas y las tecnologías más avanzadas para llevarlo a cabo.

En primer lugar, se presentan las técnicas de **Almacenes de Datos y OLAP** para facilitar el procesamiento analítico de apoyo a la toma de decisiones estratégicas.

En segundo lugar, y con más extensión, se presenta la necesidad del análisis inteligente y automático de la información para el descubrimiento de conocimiento útil.

Para abordarlo, se introducirá el área de la **Extracción de Conocimiento a partir de Bases de Datos (KDD)**, sus fases, en especial la de *minería de datos*, y se presentarán las técnicas de aprendizaje automático más habituales en minería de datos. Se estudiará la idoneidad de cada una para diferentes problemas.

El curso intercala la presentación de conocimientos teóricos con la aplicación práctica de los mismos sobre un paquete de minería de datos (IBM SPSS e IBM Info Sphere Warehouse con Intelligent Miner).

### 2. CAPACIDADES

- Familiarizarse con elementos conceptuales del marco teórico referencial de las diferentes tecnologías de los almacenes de datos, sus características y posibilidades de uso, incluyendo la utilización de los mismos con técnicas de OLAP y de minería de datos, tanto con modelos descriptivos como predictivos.
- Utilizar Almacenes de Datos, OLAP y Minería de Datos como soporte para el procesamiento de datos masivos, la extracción de conocimiento y el mejoramiento en la toma de decisiones.

### 3. CONTENIDOS CURRICULARES

**Tema 1: De las Bases de Datos a los Almacenes de Datos:** Finalidades y Evolución de los Sistemas de Información. Herramientas para la Toma de Decisiones. Estructuras de Datos para la Inteligencia de Negocios. Almacenes de Datos (data warehouses y data marts), OLAP (Proceso Analítico en Línea) y Minería De Datos.



**Tema 2: Almacenes de Datos para Gestión de Datos Masivos:** Introducción a los Almacenes de Datos. Arquitectura de un Sistema de Almacén de Datos. Carga y Mantenimiento de un Almacén de Datos. Diseño y Armado de un Almacén de Datos.

**Tema 3: OLAP:** Explotación de un Almacén de Datos: Herramientas OLAP. Sistemas ROLAP Y MOLAP. SQL Ampliado.

**Tema 4: Minería de Datos - Introducción:** Introducción a la Minería de Datos. El Proceso de KDD. Técnicas de Minería de Datos y Principales Algoritmos. Fase de Selección en Minería de Datos. Fase de Exploración en Minería de Datos. Fase de Limpieza y Transformación de Datos.

**Tema 5: Minería de Datos – Técnicas Predictivas de Modelización:** Técnicas de Minería de Datos. Técnicas Predictivas para la Modelización. Modelo de Regresión Múltiple. Modelos de Elección Discreta. Clasificación Ad Hoc: Análisis Discriminante.

**Tema 6: Minería de Datos – Técnicas Descriptivas y Predictivas de Clasificación:** El Análisis Cluster Como Técnica Descriptiva de Clasificación. Los Árboles de Decisión Como Técnica Predictiva de Clasificación. Análisis de Conglomerados y Árboles de Decisión Como Métodos de Segmentación.

**Tema 7: Minería de Datos – Redes Neuronales:** Descripción de una Red Neuronal (RN). RN y Ajuste de Modelos de Regresión. Aprendizaje en las RN. Funcionamiento de una RN. El Algoritmo de Aprendizaje de Retro-propagación. Análisis Discriminante a Través del Perceptrón. Análisis de Series Temporales Mediante RN. Análisis de Componentes Principales con RN. Agrupamiento (Clustering) Mediante RN.

**Tema 8: Minería de Datos – Minería Web:** Minería Web. Técnicas y Principales Algoritmos.

#### **4. ESTRATEGIAS DE ENSEÑANZA – APRENDIZAJE**

Se planifican las siguientes actividades: exposiciones del docente mediante clases magistrales dialogadas (con instancias de participación de los alumnos), trabajos grupales de los alumnos, estudio independiente, resolución de ejercicios de laboratorio y preparación y presentación de un trabajo final. Se utilizarán elementos de la pedagogía problematizadora, poniendo énfasis en la *motivación* para el aprendizaje de los distintos temas. Se propiciará el aprendizaje autónomo mediante la búsqueda de información en repositorios digitales de publicaciones académicas especializadas.

#### **5. CRITERIOS DE EVALUACIÓN Y APROBACIÓN**

La evaluación del proceso de enseñanza – aprendizaje se efectuará a través de instancias de consulta en las distintas clases; el trabajo final constituirá la instancia de acreditación, es decir el modo de comprobar si los alumnos han logrado los aprendizajes necesarios para aprobar el curso. En el trabajo final el alumno deberá acreditar capacidad en el manejo de los conceptos teóricos, la integración de los conocimientos, y la aplicación de los mismos en la interpretación de situaciones concretas. Se aplicará la escala de calificaciones vigente en la UNP.

## 6. BIBLIOGRAFÍA:

### BÁSICA

1. J. HERNÁNDEZ ORALLO, M. J. RAMÍREZ, C. FERRI. *Introducción a la Minería de Datos*. Pearson. 2004. ISBN 84-205-4091-9.
2. C. PÉREZ LÓPEZ, D. SANTÍN GONZÁLEZ. *Minería de Datos: Técnicas y Herramientas*. Thomson Paraninfo S. A. España. 2007. ISBN 978-84-9732-492-2.

### COMPLEMENTARIA

1. C. BALLARD, J. ROLLINS, J. RAMOS, A. PERKINS, R. HALE, A. DOERNEICH, E. CAS MILNER, J. CHODAGAM. *Dynamic Warehousing: Data Mining Made Easy*. IBM Press. USA. 2007. ISBN 0738488860.
2. O. MAIMON, L. ROKACH. *Data Mining and Knowledge Discovery Handbook*. 2/E. Springer. USA. 2010. ISBN 978-0-387-09822-7.
3. C. PÉREZ LÓPEZ. *Técnicas de Análisis de Datos con SPSS*. Pearson Prentice Hall. España. 2009. ISBN 978-84-8322-601-8.

### WEB

1. <http://users.dsic.upv.es/~jorallo/cursoDWDWM/>: J. Hernández Orallo. *Análisis y Extracción de Conocimiento en Sistemas de Información: Data Warehouse y Data Mining*. Universidad Politécnica de Valencia. España.
2. <http://www.cs.waikato.ac.nz/ml/weka/>: Weka. The University of Waikato.
3. <https://www.ibm.com/developerworks/university/academicinitiative/>: IBM. Iniciativa Académica.
4. <http://www.microstrategy.com/software/businessintelligence/>: MicroStrategy.
5. <http://www.olapcouncil.org/>: OLAP Council.
6. <http://www.spss.com/>: SPSS.