

MINERÍA DE DATOS

UC	HT	HP	HL	Modalidad	Código	Requisitos	Ult. Actualización
5	4	2		Optativa/ electiva	6213	- Bases de Datos - Probabilidad y Estadística	Junio 2004

Fundamentación:

Los progresos en la adquisición y almacenamiento de datos en diferentes ambientes, arrojan como resultado el surgimiento de enormes Bases de datos. La necesidad de explorar estas bases de datos y extraer información y conocimiento que sea de interés para los propietarios de las mismas, se ha incrementado en la actualidad. El análisis y exploración de datos se ha venido realizando con técnicas estadísticas tradicionales; esta modalidad de análisis resulta ser lenta, costosa y altamente subjetiva cuando los volúmenes de datos son muy grandes. Las técnicas de Minería de Datos han resultado ser una alternativa para solventar la búsqueda de conocimiento en bases de datos con estas características.

Objetivos:

Al finalizar el curso el estudiante deberá ser capaz de:

- Comprender los fundamentos teóricos y prácticos asociados al proceso de minería de datos.
- Comprender y aplicar técnicas de minería de datos para descubrir conocimiento en problemas del mundo real

Contenido temático:

1. Introducción al descubrimiento de conocimiento a partir de datos (Knowledge Discovery, Data-KDD).
 Antecedentes. ¿Qué es Descubrimiento de conocimiento? ¿Qué es minería de datos (MD)?
 Tareas de la MD. Componentes de los algoritmos de Minería de datos. Taxonomía de las técnicas de Minería de Datos.
2. Proceso de desarrollo de Minería de Datos.
 Pasos del proceso de MD. El proceso CRISP-DM para MD.
3. Análisis exploratorio de datos.
 ¿Que es Análisis Exploratorio de Datos? El rol de los gráficos. Técnicas de visualización para análisis de datos. Uso de las técnicas de visualización.
4. Pre-procesamiento de los Datos.
 Técnicas de Normalización. Técnicas para tratamiento de datos ausentes. Reducción de dimensionalidad
5. Algoritmos de Minería de Datos:
Algoritmos de Agrupamiento (Clustering). ¿En qué consiste el agrupamiento? ¿Qué es clasificación?. Medidas de similitud. Técnicas de clasificación. *Reglas de Asociación y Reglas de Inducción*. Introducción. Componentes de las RA. Extracción de reglas por frecuencia. *Árboles de decisión*. Árboles de decisión para clasificación. Mecanismos de Partición. Algoritmos que implantan árboles de decisión. Ejemplo ID3.
6. Aplicaciones de Minería de Datos.
 Discusión de aplicaciones de algoritmos de Minería de Datos a problemas

Bibliografía.

- Berthold, M. y Hand, D. (Eds). *Intelligent Data Analysis. An Introduction*. Springer-Verlag Berlin. 1999
- Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. *From Data Mining to Knowledge Discovery in Databases*. AI Magazing. 1996.
- Hand, D., Mannila, H. y Smyth, P. *Principles of Data Mining*. MIT Press. 2001.
- Hastie, T., Tibshirani, R. y Friedman, J. *The elements of statistical learning. Data mining, inference and prediction*. Springer USA. 2001.
- Gamberger D, Smuc, T. y Mari, I. *Data Mining Server*. Laboratory Information System. Rudjer Boskovic Institute. http://dms.irb.hr/tutorial/tut_intro.php. 2001
- Engineering Statistics. *Handbook of Statistical Method. Chapter 1: Exploratory Data Analysis*. <http://www.itl.nist.gov/div898/handbook/index.htm>