



MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

Área: Bases de Datos y Recuperación de Información

Programa de Asignatura: Minería de Datos

Código: MCOM 21100

Tipo: Obligatoria

Créditos: 9

Fecha: Noviembre 2012



1. DATOS GENERALES

Nombre del Programa Educativo:	Maestría en Ciencias de la Computación
Modalidad Académica:	Escolarizada
Nombre de la Asignatura:	Minería de Datos (Obligatoria)
Ubicación:	Segundo semestre

2. REVISIONES Y ACTUALIZACIONES

Autores:	Dra. María Josefa Somodevilla García
Fecha de diseño:	Noviembre 2012
Fecha de la última actualización:	Marzo 2017
Revisores:	Dr. Ivo Humberto Pineda Torres Dra. Darnes Vilariño Ayala Dr. David Eduardo Pinto Avendaño Dra. María de la Concepción Pérez de Celis Herrero Dra. María Josefa Somodevilla García
Sinopsis de la revisión y/o actualización:	Con base en los resultados de la evaluación del programa de asignatura por los actores (estudiantes, profesor y academia) se determinó que el programa precedente de la materia no era suficiente en términos de tiempo. Se incorporaron dos capítulos: Minería de Grafos de Redes Sociales y Máquinas de Aprendizaje a gran escala. Estos nuevos temas extienden el programa para considerar las necesidades actuales de generación de conocimiento a través de la minería de conjuntos de datos masivos.



3. OBJETIVOS:

1. **General:** Distinguir los sistemas de información para la gestión de los sistemas de información para la toma de decisiones, determinar la necesidad de desarrollo de almacenes de datos y reconocer tipos de problemas de minería de datos así como los métodos adecuados de acuerdo al conjunto de datos a ser aplicados.

Específicos:

1. Conocer la evolución de las herramientas para el análisis de sistemas de información para la toma de decisiones.
2. Conocer las ventajas y casos donde es aconsejable recopilar información interna y externa en un Almacén de Datos.
3. Conocer el modelo multidimensional de los almacenes de datos y los operadores de refinamiento asociados: *drill, roll, slice & dice, pivot*.
4. Conocer la arquitectura y diferentes implementaciones (ROLAP, MOLAP) de Almacenes de Datos.
5. Reconocer pautas para el diseño y mantenimiento de ADs.
6. Reconocer la problemática del análisis de grandes volúmenes de datos y de los beneficios de su uso sistemático para la obtención de modelos y patrones predictivos o descriptivos.
7. Conocer las fases del Descubrimiento de Conocimiento de Bases de Datos y la importancia de las mismas en el éxito del proceso (en especial las de limpieza y selección de datos).
2. Conocer las distintas técnicas de aprendizaje automático y estadísticas utilizadas en minería de datos, su potencial, su coste computacional y sus limitaciones de representación y de inteligibilidad.
3. Elegir, para un problema concreto, qué técnicas de minería de datos son más apropiadas.
4. Generar los modelos y patrones elegidos utilizando una herramienta o paquete de minería de datos.
5. Evaluar la calidad de un modelo, utilizando técnicas sencillas de evaluación (validación cruzada).



4. CONTENIDO

Unidad	Contenido Temático	Bibliografía	
		Básica	Complementaria
1. Introducción	1.1. Finalidad y Evolución de los Sistemas de Información. 1.2. Herramientas para la toma de decisiones: diferencias e interrelación. 1.3. Almacenes de Datos, OLAP y Minería de Datos: definición e interrelación.	Hernández Orallo et al., <i>"Introducción a la Minería de Datos"</i> . Pearson, Prentice Hall, 2004	Pang-Ning Tan, Michael Steinbach y Vipin Kumar, <i>"Introduction to Data Mining"</i> . Addison Wesley. 2005
2. Almacenes de Datos	2.1. Introducción a los Almacenes de Datos. 2.2. Arquitectura de un Sistema de Almacén de Datos. 2.3. Explotación de un Almacén de Datos: Herramientas OLAP. 2.4. Sistemas ROLAP y MOLAP. 2.5. Carga y Mantenimiento de un Almacén de Datos. 2.6. Diseño de un Almacén de Datos.	Kimball R. and Ross M, <i>"The Data Warehouse Lifecycle Toolkit"</i> , Wiley Publishing Inc., 2008	Hernández Orallo et al., <i>"Introducción a la Minería de Datos"</i> . Pearson, Prentice Hall. 2004
3. Descubrimiento de Conocimiento en Bases de Datos	3.1. Introducción a la Minería de Datos (DM) 3.2. Preparación de Datos (proceso KDD) 3.3. El Problema de la Extracción Automática de Conocimiento. 3.4. Evaluación de Hipótesis	Ian H. Witten and Eibe Frank , <i>"Data Mining: Practical Machine Learning Tools and Techniques"</i> Third Edition, 2011	Pérez Cesar y Santín Daniel. <i>"Data Mining: Técnicas y Herramientas"</i> . Alfa Omega, 2007
4. Técnicas de Minería de	4.1. Técnicas no supervisadas y descriptivas. 4.1.1 Análisis	Ian H. Witten and Eibe Frank, <i>"Data Mining: Practical"</i>	Jiawei Han y Micheline Kamber, <i>"Data Mining: Concepts"</i>



Unidad	Contenido Temático	Bibliografía	
		Básica	Complementaria
Datos	Exploratorio 4.1.2. Regresión 4.1.3 Reglas de Asociación 4.1.4 Agrupamiento 4.2. Técnicas supervisadas y predictivas. 4.2.1 Predicción 4.2.2 Redes Neuronales 4.2.3 Árboles de Decisión 4.2.3 Clasificadores Bayesianos	<i>Machine Learning Tools and Techniques</i> Third Edition. 2011	<i>and Techniques</i> . Morgan Kaufmann Publishers. Second Edition 2006 http://www.cs.waikato.ac.nz/~ml/index.html (Página principal de WEKA) DATASETS http://archive.ics.uci.edu/ml/
5. Minería de Grafos de Redes Sociales	5.1 Redes Sociales como grafos 5.2 Agrupamiento de grafos de Redes Sociales 5.3 Particionamiento de grafos 5.4 Propiedades de vecindario de grafos	Jure Leskovec, Anand Rajaraman Jeffrey D. Ullman "Mining of Massive Datasets" Second Edition 2014	By Matthew A. Russell "Mining the Social Web" Orelly 2nd Edition 2013
6. Máquinas de Aprendizaje a gran escala	6.1 El modelo de aprendizaje automático 6.2 Perceptrones 6.3 Máquinas de Vector de Soporte 6.4 Vecinos más cercanos 6.5 Comparación de métodos de aprendizaje	Jure Leskovec, Anand Rajaraman Jeffrey D. Ullman "Mining of Massive Datasets" Second Edition 2014	Jared Dean "Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners" 2014



5. CRITERIOS DE EVALUACIÓN

CRITERIO	PORCENTAJE (%)
Exámenes Parciales (2)	40
Tareas	20
Proyecto de Investigación	20
Escritura de Artículo de Investigación	20
TOTAL	100