



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:	Abril 2019
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 debería considerar nuevos paradigmas computacionales para generar soluciones en el área de descubrimiento de conocimiento. Se modificaron los objetivos y se actualizó la bibliografía de manera que se considerarán las necesidades actuales de generación de conocimiento a través de la minería de conjuntos de datos masivos.



3. OBJETIVOS:

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 el modelo multidimensional de los almacenes de datos y los operadores de refinamiento asociados, así como su arquitectura y diferentes implementaciones.
3. 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.
4. 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).
5. Conocer las distintas técnicas de aprendizaje automático utilizadas en minería de datos, su potencial, su coste computacional y sus limitaciones de representación y de inteligibilidad.
6. Elegir, para un problema concreto, qué técnicas de minería de datos son más apropiadas.
7. Generar los modelos y patrones elegidos utilizando una herramienta de aprendizaje automático.
8. Evaluar la calidad de un modelo, utilizando técnicas de evaluación adecuadas para el conjunto de datos de aprendizaje.



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. (2004). Introducción a la Minería de Datos. USA: Pearson, Prentice Hall.	Pang-Ning Tan, Michael Steinbach y Vipin Kumar (2019). Introduction to Data Mining, 2nd edition. USA: Addison Wesley.
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.	Ralph Kimball and Margy Ross. (2014). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition. USA: John Wley & Sons.	Hernández Orallo et al. (2004). Introducción a la Minería de Datos. USA: Pearson, Prentice Hall.
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. (2017). Practical Machine Learning Tools and Techniques 4th Edition. USA: Morgan Kufmann.	Hernández Orallo et al. (2004). Introducción a la Minería de Datos. USA: Pearson, Prentice Hall.
4. Técnicas de Minería de Datos	4.1. Técnicas no supervisadas y descriptivas. 4.1.1 Análisis Exploratorio 4.1.2. Regresión	Charu C Aggarwal. (2015). Data Mining: The Textbook. USA: Springer.	Jiawei Han y Micheline Kamber. (2012). Data Minig: Concepts and Techniques, 3th



Unidad	Contenido Temático	Bibliografía	
		Básica	Complementaria
	4.1.3 Reglas de Asociación 4.1.4 Agrupamiento 4.2. Técnicas supervisadas y predictivas. 4.2.1 Predicción 4.2.2 RedesNeuronales 4.2.3 Árboles de Decisión 4.2.3 Clasificadores Bayesianos		edition. USA: Morgan Kaufmann. Python Machine Learning: Scikit-Learn Tutorial. (2019). Python Machine Learning: Scikit-Learn Tutorial. 18/04/2019, de DataCamp Sitio web: https://www.datacamp.com/community/tutorials/machine-learning-python
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. (2014). Mining of Massive Datasets Second Edition. USA: Cambridge University Press.	Matthew A. Russell. (2019). Mining the Social Web, 3th edition. USA: Orelly.
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	John Hearty. (2017). Large Scale Machine Learning With Python. USA: Books Library Land.	Jared Dean. (2014). Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners. USA: Wiley.



5. CRITERIOS DE EVALUACIÓN

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