



PROGRAMA DE ESTUDIO POR COMPETENCIAS  
Minería de Datos

I. IDENTIFICACIÓN DEL CURSO

<b>ORGANISMO ACADÉMICO: FACULTAD DE INGENIERÍA</b>						
<b>Programa Educativo:</b> Ingeniería en Computación				<b>Área de docencia:</b> Tratamiento de la Información		
<b>Aprobación por los H.H. Consejos Académico y de Gobierno</b>		<b>Fecha:</b>		<b>Programa elaborado por:</b> M. en I. Patricia Romero Ramírez		<b>Programa revisado por:</b> Centro Universitario de Chalco Centro Universitario Texcoco Centro Universitario del Valle de México M en I. Patricia Romero Ramírez
				<b>Fecha de elaboración :</b> 19/12/06		<b>Fecha de revisión :</b> Mayo de 2010
<b>Clave</b>	<b>Horas de teoría</b>	<b>Horas de práctica</b>	<b>Total de horas</b>	<b>Créditos</b>	<b>Tipo de Unidad de Aprendizaje</b>	<b>Núcleo de formación</b>
L41093	2	2	4	6	CURSO	Integral
<b>Prerrequisitos:</b> Fundamentos de bases de datos. Sistemas gestores de bases de datos, bases de datos avanzadas		<b>Unidad de Aprendizaje Antecedente:</b> Ninguna		<b>Unidad de Aprendizaje Consecuente:</b> Ninguna		
<b>Programas educativos o espacios académicos en los que se imparte:</b> Licenciatura en Ingeniería en Computación (Facultad. de Ingeniería, Centros Universitarios: Atlacomulco, Ecatepec, Texcoco, Valle de Chalco, Valle de México, Valle de Teotihuacán, Zumpango)						



## **II. PRESENTACIÓN DEL PROGRAMA**

En la actualidad, el proceso de extracción de información útil requiere la aplicación de técnicas de análisis de datos automáticas que sean capaces de procesar grandes volúmenes de información. La minería de datos es el proceso de aplicar metodologías basadas en computadores, conceptos estadísticos y administración para extraer conocimiento útil de grandes volúmenes de información.

El propósito de este curso es estudiar y aplicar el proceso metodológico, las bases teóricas de las técnicas de minería más usuales y las herramientas de software disponibles para poder realizar un proyecto de minería de datos en un contexto real.

En la primera unidad se estudian los conceptos necesarios como bases de esta asignatura con respecto a bases de datos, probabilidad y estadística, Almacenes de datos, OLTP vs OLAP, definición de Minería de Datos, relación con otras disciplinas y ejemplos de aplicación.

En la segunda unidad se describe el proceso de extracción de conocimiento KDD (Knowledge Discovery in Databases) explicando cada una de las etapas del mismo: Integración y recopilación; Selección, limpieza y transformación; Minerías de Datos; Evaluación e Interpretación; Difusión y uso.

La tercera unidad profundiza el estudio de los métodos más usuales que permiten encontrar patrones en la información para realizar tareas tanto predictivas (clasificación, regresión y agrupamiento) como descriptivas (agrupamiento, reglas de asociación y correlaciones). Las técnicas o métodos que se estudian son: redes neuronales, arboles de decisión, regresión, kmeans, reglas de asociación, Naive Bayes, vecinos más próximos, algoritmos genéticos y evolutivos entre otras.

La cuarta unidad se enfoca al conocimiento del software disponible para aplicar las técnicas estudiadas sobre la vista minable obtenida. Este estudio se realiza en tres enfoques: API para minería, SQL extendido para minería y ambientes visuales. Algunas de las herramientas disponibles actualmente son: Weka, Clementine, Darwin, Rapid Miner y SQL server para minería entre otros.

Finalmente en la quinta unidad se estudia la metodología CRISP-DM y las tendencias en minería de datos textuales, secuenciales, espaciales, multimedia, temporales y Web.

Para consolidar el aprendizaje es necesario realizar prácticas en el software disponible así como también la realización de un proyecto en un contexto real en el que sean aplicados los conocimientos adquiridos involucrando todo el proceso metodológico desde la comprensión de los objetivos del negocio hasta la obtención y evaluación de resultados. Como parte de la investigación por parte de los alumnos se sugiere revisar sitios web de proyectos de minería actuales por ejemplo: SAS Text Miner, WebDataKit, IBM Intelligent Miner for Text, OK Log, etc.

La evaluación de la unidad de aprendizaje se efectuará a través de dos evaluaciones parciales una de las cuales puede ser suplida por la exposición por parte de los alumnos de las técnicas de minería estudiadas con ejemplos de aplicación en el software disponible.

Dada la importancia de esta asignatura en la vida profesional actual se recomienda realizar video filmación de los proyectos realizados en cada curso lo cual permitirá contar con un histórico de ejemplos reales logrados por los alumnos que serán de utilidad en cursos posteriores como material digital de enseñanza.

Con lo antes mencionado, el ingeniero en computación como experto en su ramo tendrá las herramientas necesarias para poder interactuar con profesionales en otros campos del saber para así aplicar minería de datos con bases científico-metodológicas que le permitan afrontar los retos actuales de la extracción de conocimientos a partir de grandes repositorios de información.



### III. LINEAMIENTOS DE LA UNIDAD DE APRENDIZAJE

<b>DEL DOCENTE</b>	<b>DEL DISCENTE :</b>
<ul style="list-style-type: none"><li>• Establecer las políticas del curso.</li><li>• Respetar el horario del curso y la forma de evaluarlo.</li><li>• Cumplir el temario y el número de horas asignadas al curso.</li><li>• Asesorar y guiar el trabajo de las unidades de aprendizaje.</li><li>• Retroalimentar el trabajo de los alumnos.</li><li>• Fomentar la creatividad en los alumnos a través del desarrollo de proyectos.</li><li>• Elaborar y estar presente en las prácticas.</li></ul>	<ul style="list-style-type: none"><li>• Contar con la asistencia establecida en el reglamento de Facultades:</li><li>• 80% para examen ordinario</li><li>• 60% para examen extraordinario</li><li>• 30% para examen a título de suficiencia</li><li>• Cumplir con las actividades encomendadas entregando con calidad en tiempo y forma los trabajos requeridos</li><li>• Participar activa y críticamente en el proceso de enseñanza-aprendizaje</li><li>• Hacer uso adecuado de las instalaciones y equipo de cómputo</li></ul>

### IV. PROPÓSITO DE LA UNIDAD DE APRENDIZAJE

El alumno: Estudiara y aplicara el proceso metodológico, las bases teóricas de las técnicas de minería de datos más usuales y las herramientas de software disponibles para adquirir la capacidad de aplicar minería de datos en un contexto real.



## **V. COMPETENCIAS GENÉRICAS**

El alumno:

- Conocerá las bases metodológicas para la extracción de conocimiento a partir de grandes repositorios de información.
- Conocerá las bases teóricas de las técnicas de minería más usuales que permiten la obtención de modelos y patrones a partir de la vista minable.
- Conocerá el software disponible para realizar minería de datos.
- Podrá conjuntar los conocimientos adquiridos para realizar un proyecto de minería de datos en un contexto real.
- Podrá interpretar y evaluar los resultados obtenidos para trasladarlos en acciones recomendadas.
- Conocerá la metodología CRISP-DM y la minería de datos complejos.

## **VI. ÁMBITOS DE DESEMPEÑO**

- Empresas públicas y privadas que manejan grandes volúmenes de información.
- Consultoría en proyectos de minería de datos.

## **VII. ESCENARIOS DE APRENDIZAJE**

- Salón de Clases
- Sala de Cómputo



**VIII. ESTRUCTURA DE LA UNIDAD DE APRENDIZAJE**

- Introducción a la minería de datos
- La minería de datos en el proceso de KDD (Knowledge Discovery in Databases)
- Tareas y métodos en minería de datos
- Herramientas de software disponibles para minería de datos
- Metodología CRISP-DM y minería de datos complejos



IX. DESARROLLO DE LA UNIDAD DE APRENDIZAJE

UNIDAD DE COMPETENCIA I	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes/ Valores
Introducción a la minería de datos	Conceptos básicos Disciplinas involucradas Definición de minería de datos Almacenes de datos (OLAP vs OLTP) Aplicaciones de la minería de datos	Observación Conceptualización Análisis	Responsabilidad para cumplir con las tareas asignadas Actitud propositiva, constructivista e innovadora Tolerancia y participación Trabajo en equipo Receptiva Analítica
<b>ESTRATEGIAS DIDÁCTICAS:</b> Revisión y análisis de textos Explicación por parte del instructor de temas específicos de mayor complejidad	<b>RECURSOS REQUERIDOS</b> - Pizarrón de apoyo - Computadora Diapositivas o videos	- Libros - Cañón	<b>TIEMPO DESTINADO</b> 4 horas teóricas 4 horas prácticas
<b>CRITERIOS DE DESEMPEÑO I</b>	<b>EVIDENCIAS</b>		
	<b>DESEMPEÑO</b>	<b>PRODUCTOS</b>	
Investigará que es la minería de datos y aplicaciones	Discusión dirigida Reporte de la investigación.	Definir y comprender el campo de la minería de datos	
Recordará temas de otras áreas de conocimiento base de la minería de datos.	Realización de ejercicios básicos	Recordar conocimientos de base para comprender el campo de la minería de datos.	
Resumirá las disciplinas involucradas, aplicaciones y enfoques de la minería de datos	Cuadro sinóptico o mapa conceptual	Comprender el alcance de la minería de datos.	



UNIDAD DE COMPETENCIA II	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes/ Valores
La minería de datos en el proceso de KDD (Knowledge Discovery in Databases)	<b>Etapas de proceso de KDD:</b> Integración y recopilación; Selección, limpieza y transformación; Minería de Datos; Evaluación e Interpretación; Difusión y uso.	Comprensión del proceso de extracción de conocimiento del cual forma parte la minería de datos.	Responsabilidad para cumplir con las tareas asignadas Actitud propositiva, constructivista e innovadora Tolerancia y participación Trabajo en equipo Receptiva  Analítica
<b>ESTRATEGIAS DIDÁCTICAS:</b> Revisión y análisis de textos Explicación por parte del instructor de temas específicos de mayor complejidad Solución de ejercicios		<b>RECURSOS REQUERIDOS</b> - Pizarrón - Computadora - Diapositivas o videos - Cañón	<b>TIEMPO DESTINADO</b>  4 horas teóricas 2 horas prácticas
<b>CRITERIOS DE DESEMPEÑO II</b>	<b>EVIDENCIAS</b>		
		<b>DESEMPEÑO</b>	<b>PRODUCTOS</b>
1. Conocer y aplicar las fases del KDD		Conocer el proceso de KDD	Obtener vistas minables a partir de ejemplos
2. Elegir, para un caso de aplicación ya desarrollado e identificar las fases del proceso de KDD		Ejercicio practico	Identificar adecuadamente las fases de KDD.



UNIDAD DE COMPETENCIA III	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes/ Valores
Tareas y métodos en minería de datos.	Tareas predictivas (clasificación, regresión y agrupamiento). Tareas descriptivas (agrupamiento, reglas de asociación y correlaciones). Las técnicas o métodos que se estudian son: redes neuronales, arboles de decisión, regresión, kmeans, reglas de asociación, Naive Bayes, vecinos más próximos, algoritmos genéticos y evolutivos entre otras.	Conceptualización Análisis de bases teóricas Creatividad	Responsabilidad para cumplir con las tareas asignadas Actitud propositiva, constructivista e innovadora Tolerancia y participación Trabajo en equipo Receptiva Analítica
<b>ESTRATEGIAS DIDÁCTICAS:</b> Revisión y análisis de textos Explicación por parte del instructor y/o discente de temas específicos		<b>RECURSOS REQUERIDOS</b> - Pizarrón - Cañón - Computadora - Diapositivas o videos	<b>TIEMPO DESTINADO</b>  4 horas teóricas 2 horas prácticas
<b>CRITERIOS DE DESEMPEÑO III</b>	<b>EVIDENCIAS</b>		
		<b>DESEMPEÑO</b>	<b>PRODUCTOS</b>
Aplicar las bases teóricas de las técnicas estudiadas en ejemplos de aplicación.		Identificación de las situaciones en las que conviene la aplicación de cada técnica estudiada	Ejemplos resueltos con diversas técnicas.



UNIDAD DE COMPETENCIA IV	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes/ Valores
Herramientas de software disponibles para minería de datos	API's para integración dentro de código fuente (JDM, WEKA, etc.) Lenguajes de consulta (M-SQL, DMQL, Darwin, etc.) Entornos visuales (Clementine, Weka, Rapid Miner, etc.)	Conceptualización Análisis Creatividad	Responsabilidad para cumplir con las tareas asignadas Actitud propositiva, constructivista e innovadora Tolerancia y participación Trabajo en equipo Receptiva Analítica
<b>ESTRATEGIAS DIDÁCTICAS:</b> Buscar, descargar y utilizar el software disponible para minería de datos. Talleres de práctica del uso del software a utilizar.		<b>RECURSOS REQUERIDOS</b> - Pizarrón - Libro de texto - Cañón - Diapositivas o videos Computadora	<b>TIEMPO DESTINADO</b> 8 horas teóricas 8 horas prácticas
CRITERIOS DE DESEMPEÑO IV	EVIDENCIAS		
	DESEMPEÑO	PRODUCTOS	
1. Generar los modelos y patrones utilizando las herramientas de minería de datos disponibles	Aplicación de las técnicas estudiadas en ejemplos	Conocer las distintas técnicas utilizadas en minería de datos.	
2. Aplicación del software disponible al proyecto del curso.	Revisión de avances de proyectos	Aplicar las técnicas estudiadas.	



UNIDAD DE COMPETENCIA V	ELEMENTOS DE COMPETENCIA		
	Conocimientos	Habilidades	Actitudes/ Valores
Metodología CRIPS-DM y minería de datos complejos	Metodología CRIPS-DM Ciclo en espiral para proyectos de minería.  Minería de datos textuales, secuenciales, espaciales, temporales, multimedia y Web	Conceptualización Análisis Creatividad	Responsabilidad para cumplir con las tareas asignadas Actitud propositiva, constructivista e innovadora Tolerancia y participación Trabajo en equipo Receptiva Analítica
<b>ESTRATEGIAS DIDÁCTICAS:</b> Revisión y análisis de textos Explicación por parte del instructor y/o discente de temas específicos		<b>RECURSOS REQUERIDOS</b> - Pizarrón - Libros de apoyo - Cañón - Diapositivas o videos - Computadora	<b>TIEMPO DESTINADO</b>  4 horas teóricas 8 horas prácticas
<b>CRITERIOS DE DESEMPEÑO IV</b>	<b>EVIDENCIAS</b>		
		<b>DESEMPEÑO</b>	<b>PRODUCTOS</b>
1. Complementar las bases metodológicas para aplicar minería de datos en un contexto real.		Aplicación de la metodología CRISP-DM en el proyecto del curso	Documentación del proyecto en base a la metodología estudiada
2. Comprender la minería de datos complejos.		Investigación de sitios relacionados a la minería de datos complejos	Participación activa en la búsqueda de información



## X. EVALUACIÓN Y ACREDITACIÓN

### Ordinario:

2 exámenes parciales	40%
Tareas y trabajo en clase	10%
Proyecto final	20%
Examen final	<u>30%</u>
	100%

El alumno podrá exentar si en los tres primeros rubros en escala de 0 – 70, obtiene sobre estos rubros un porcentaje  $\geq 80$ . Debido a que es evaluación continua, al alumno se le asentará una calificación numérica, la cual obtiene durante el curso.

### Extraordinario:

Examen escrito	60%
Proyecto final	<u>40%</u>
	100%

### Título de suficiencia:

Examen escrito	60%
Proyecto final	<u>40%</u>
	100%

### Acreditación:

- Para tener derecho a cualquier examen deberá entregar y obtener calificación aprobatoria en el proyecto final de aplicación.
- Asistencia: Ordinario: asistencia  $\geq 80\%$   
Extraordinario: asistencia  $\geq 60\%$   
A título de suficiencia: asistencia  $\geq 30\%$   
Sin derecho a examen:  $30\% >$  asistencia



## **XI. BIBLIOGRAFÍA**

### **Bibliografía básica**

- Hernández Orallo, J., M. J. Ramírez Quintana, et al. (2004). Introducción a la Minería de Datos. España, Pearson Educación SA.
- Han, D. J. (2007). Principles of Data Mining, MIT Press.
- Maimon, O. Z. and L. Rokach (2005). Data mining and knowledge discovery handbook. USA, Springer.
- Pérez López, C. and D. Santín Gonzalez (2006). Data Mining-Soluciones Con Enterprise Miner. México, Alfaomega, Ra-Ma.
- Sumathi, S. and S. N. Sivanandam (2006). Introduction to data mining and its applications. Berlín, Germany, Springer-Verlag New York Inc.
- Tan, P. N., M. Steinbach, et al. (2005). Introduction to data mining, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

### **Bibliografía complementaria**

- Everitt, B.S. (1994). A Handbook of Statistical Analyses using S-Plus. Chapman and Hall.
- Inmon, W.H. (1996). Building the Datawarehouse. J.Wiley & Sons.
- Han, J. and M. Kamber (2006). Data mining: concepts and techniques, Morgan Kaufmann.
- Kimball, R (1996). The Data Warehouse Toolkit. John Wiley & Sons.
- Hastie, T., R. Tibshirani, et al. (2005). The elements of statistical learning: data mining, inference and prediction, Springer
- Dunham. H. Margaret (2003). Data Mining. Introductory and Advanced Topics, Prentice Hall.
- Pyle, D. (1999), "Data Preparation for Data Mining", Morgan Kaufmann, San Francisco, CA.
- Hand, David; Mannila, Heikki; Smyth, Padhraic (2001), Principles of Data Mining, A Bradford Book. The MIT Press.
- Ian Witten and Eibe Frank (2002), Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers.