

PROGRAMA DE CURSO

Código	Nombre			
IN 4521	INTRODUCCIÓN A LA MINERÍA DE DATOS			
Nombre en Inglés				
Introduction to Data Mining				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	10	3	1,5	5,5
Requisitos			Carácter del Curso	
MA3403 Probabilidades y Estadísticas			Electivo para la Carrera de Ingeniería Civil Industrial	
Resultados de Aprendizaje				
Al finalizar el curso, el alumno:				
1. Conoce modelos de Bases de Datos, Bodegas de Datos, OLAP, entendiendo sus usos y en aplicaciones en las organizaciones. 2. Entiende los principales objetivos y métodos de Data Mining. 3. Conoce aplicaciones típicas de Data Mining y tiene nociones sobre el uso de herramientas computacionales para la solución y aplicación de distintos métodos de Minería de Datos. 4. Tiene el conocimiento y la capacidad para elegir el método adecuado en una aplicación real y presentar una solución a dicha problemática.				

Metodología Docente	Evaluación General
Los alumnos, en forma adicional a las clases de cátedra, desarrollarán cuatro tareas en el semestre y CTP's. Tarea 1: Análisis multidimensional y Limpieza de datos El objetivo de esta tarea es que se entiendan los conceptos de OLAP y el uso de una herramienta para el análisis multidimensional de datos, junto con la limpieza y consolidación de datos con el fin de dar soluciones y recomendaciones a una empresa sobre las políticas comerciales a seguir. Tarea 2: Preprocesamiento, Transformación y Visualización de los datos El objetivo de esta tarea es entender la	Las tareas deberán ser realizadas en grupos de tres personas. El profesor se guarda el derecho de interrogar a cualquier integrante de un grupo para evaluar el trabajo del grupo completo, esto consiste en la elección de algún integrante al azar, al cual se le harán preguntas sobre la tarea o tendrá que mostrar manejo del software. La ponderación de la nota de cada tarea será: Nota Tarea = 0.7* Nota Informe + 0.3* Nota Interrogación. Nota Tarea Final = Promedio de las tareas. En conjunto con las tareas se realizarán CTPs. Al final del semestre todos los alumnos rendirán un control.

necesidad del preprocesamiento y la transformación de los datos y conocer algunas técnicas básicas para éste. Se trabajará con la herramienta RapidMiner. Tarea 3: Aplicación de técnicas de Minería de Datos Modelación de un problema y aplicación de diferentes técnicas del área de Minería de Datos. El objetivo de esta tarea es adquirir el conocimiento necesario para aplicar técnicas de Minería de Datos. Haciendo un análisis crítico sobre modificaciones en los parámetros de los distintos modelos, con el fin de obtener posibles mejoras en la solución. Se trabajará con la herramienta R. Tarea 4: Presentación de una solución El objetivo es que los alumnos resuelvan un problema real desde la fase de la definición del problema hasta la presentación final de la solución. La idea de esta tarea es la simulación de un proyecto real, teniendo la capacidad de elegir el mejor método para la resolución del problema.	Para aprobar el curso se necesita lo siguiente: Nota control ≥ 4.0 Nota promedio tarea ≥ 4.0 Nota CTP ≥ 5.0 La nota final del curso se determina en la siguiente manera: <table> <tr> <td>Nota control</td><td>40%</td></tr> <tr> <td>Nota promedio de tareas</td><td>40%</td></tr> <tr> <td>Nota CTP</td><td>20%</td></tr> </table>	Nota control	40%	Nota promedio de tareas	40%	Nota CTP	20%
Nota control	40%						
Nota promedio de tareas	40%						
Nota CTP	20%						

UNIDADES TEMÁTICAS

Número	Nombre de la Unidad	Duración en Semanas
1	INTRODUCCIÓN Y MOTIVACIÓN	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción al curso. 2. Presentación externa: un profesional del área presenta su experiencia trabajando con los métodos y conceptos del curso.	El alumno: Reconoce el impacto que el contenido de este curso puede tener en empresas privadas y organizaciones públicas.	

Número	Nombre de la Unidad	Duración en Semanas
2	CONCEPTOS PARA GUARDAR Y ACCEDER A LOS DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Revisión del modelo de Bases de Datos, y la evolución de las tecnologías asociadas a Bases de Datos, dando nociones básicas sobre Data Warehouse, Data Marts y OLAP.	El alumno: Es capaz de construir una base de datos simple y manejan las técnicas básicas para la manipulación de datos.	

Número	Nombre de la Unidad	Duración en Semanas
3	PROCESO KDD	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción al proceso KDD (Knowledge Discovery Databases) y su relación con la Minería de Datos. Diferentes formas de aprendizaje automático.	El alumno: Adquiere una vista general del proceso para analizar datos y encontrar información.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
4	CALIDAD DE DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Los diferentes conceptos de errores en los datos disponibles.	El alumno: Es capaz de reconocer errores en los datos y manejar las herramientas correspondientes.	

Número	Nombre de la Unidad	Duración en Semanas
5	SELECCIÓN DE DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Métodos para selección de atributos	El alumno: Reconoce la importancia de la selección de atributos. Es capaz de identificar los métodos más indicados para la selección de atributos y maneja las herramientas correspondientes.	

Número	Nombre de la Unidad	Duración en Semanas
6	TRANSFORMACIÓN DE DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Conceptos para la transformación de los datos	El alumno: Adquiere experiencia en la transformación de datos y reconoce su impacto en los resultados del posterior análisis.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
7	MÉTODOS DE LA MINERÍA DE DATOS	3.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Regresión logística, árboles de decisión, métodos de clustering, redes neuronales, Support Vector Machines.	El alumno: Es capaz de identificar el método más adecuado en una tarea dada y tiene experiencia con el manejo de herramientas computacionales.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
8	APLICACIONES	3.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Casos de segmentación de clientes, detección de fraude, análisis de riesgo, entre otros.	El alumno: Adquiere experiencia con la aplicación de la minería de datos a través del análisis de diversas aplicaciones.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
9	DIRECCIONES DE INVESTIGACIÓN RECIENTES	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Minería de datos y teoría de juegos, selección de atributos, entre otros.	El alumno: Reconoce el estado del arte actual y tiene una noción de la investigación actual del área.	

Bibliografía General

Bibliografía obligatoria:

Será publicada en UCursos.

Bibliografía complementaria:

Shmueli, G., Patel, N. R., Bruce, P. C. (2010): Data Mining for Business Intelligence. 2nd ed., John Wiley and Sons, Hoboken, New Jersey

Vigencia desde:	Otoño 2011
Elaborado por:	Richard Weber
Aprobado por:	Comisión de Docencia DII