

## PROGRAMA DE CURSO

Código	Nombre			
CC5206	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	3.0	0	7
Requisitos			Carácter del Curso	
CC3001 Algoritmos y Estructuras de Datos ó FI2002 Electromagnetismo			Electivo para ICC	
Competencia a la que Tributa el Curso				
<ul style="list-style-type: none"> <li>- Generar información útil y relevante a partir de la aplicación de técnicas de análisis de datos.</li> <li>- Encontrar asociaciones implícitas o no evidentes entre las diferentes dimensiones de los datos mediante técnicas de minería de datos.</li> <li>- Formular consultas, evaluar e interpretar adecuadamente los resultados obtenidos de la aplicación de las metodologías de análisis de datos.</li> </ul>				
Propósito del Curso				
<p>Al finalizar este curso se espera que los estudiantes generen conocimiento a partir de diversos tipos de datos, aplicando el proceso de “knowledge discovery in databases”, con énfasis en las técnicas principales de Minería de Datos. Los estudiantes tendrán oportunidad de conocer, identificar y seleccionar las técnicas de análisis que, según los criterios presentados en el curso, mejor se apliquen al objetivo de generación de conocimiento, según el dominio del problema planteado. Para esto se desarrollarán casos que incorporen problemáticas y datos de áreas tales como la Astronomía, Economía, Medicina, Marketing, entre otras. Se espera, además, acercar a los estudiantes a la problemática del análisis de grandes volúmenes de datos. También alertarlos respecto a prácticas de sobreajuste de los modelos de procesamiento de información que impiden un análisis objetivo de los datos, desarrollando con esto último, habilidades éticas de desempeño profesional en el ámbito de Ingeniería de Datos.</p>				
Resultados de Aprendizaje				
<p>Al finalizar el curso el estudiante:</p> <ol style="list-style-type: none"> <li>1. Construye y valida hipótesis de generación de información, a través de la aplicación de un conjunto de técnicas básicas de minería de datos.</li> <li>2. Interpreta los resultados del proceso de minería de datos en relación a las hipótesis planteadas, distinguiendo patrones novedosos y significativos de información.</li> </ol>				

Indicadores			
Resultado Aprendizaje	Indicadores de Desempeño	Actividades	Evaluaciones
<p>1. Construyen y validan hipótesis de generación de información, a través de la aplicación de un conjunto de técnicas básicas de minería de datos.</p>	<ul style="list-style-type: none"> <li>• Define qué es la minería de datos.</li> <li>• Señala las componentes del proceso de minería de datos.</li> <li>• Señala qué técnica de minería de datos aplicar para problemáticas simples.</li> <li>• Aplican técnicas de minería de datos en distintos análisis de casos y proyecto.</li> </ul>	<ul style="list-style-type: none"> <li>• Discusión en clases de ejemplos de casos de uso.</li> <li>• Caso con un problema definido a desarrollar en clases (1 problema para c/unidad 2, 3, 4).</li> <li>• Proyecto grupal</li> <li>• Trabajo en clase sobre proyecto grupal.</li> </ul>	<p>Presentación I + Informe del Proyecto Grupal: Datos elegidos, hipótesis iniciales y posibles técnicas (mide la construcción y validación de hipótesis)</p> <p>Actividades prácticas (casos/clases prácticas) (mide la aplicación).</p>
<p>2. Interpretan los resultados del proceso de minería de datos en relación a las hipótesis planteadas, distinguiendo patrones novedosos y significativos de información.</p>	<ul style="list-style-type: none"> <li>• Interpreta los resultados obtenidos de diferentes técnicas</li> <li>• Reformulan la estrategia para abordar el problema .</li> <li>• Contrastan/evalúan sus resultados con los indicadores establecidos para cada técnica.</li> </ul>	<ul style="list-style-type: none"> <li>• Discusión en clases de ejemplos de casos de uso.</li> <li>• Caso con un problema definido a desarrollar en clases (1 problema para c/unidad 2, 3, 4).</li> <li>• Proyecto grupal</li> <li>• Trabajo en clase sobre proyecto grupal.</li> </ul>	<p>Presentación II + Informe: Resultados del proceso de minería de datos inicial (mide interpretación + reformulación).</p> <p>Presentación III: Resultados finales.</p>

### Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas	
1	Introducción	3	
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía	
1. Introducción a la minería de datos.  2. Introducción a las técnicas principales de MD y sus aplicaciones (a. Supervisado, b. No-supervisado, Reglas de Asociación).	El estudiante: <ul style="list-style-type: none"> <li>- Define que es la minería de datos.</li> <li>- Conoce las componentes del proceso de minería de datos.</li> <li>- Conoce qué técnicas de minería de datos aplicar para problemáticas simples.</li> </ul>	[1] Cap.1, Cap.2 Cap. 3	[3] Cap. 1, Cap. 2
		[4] Cap. 1., Cap. 4.7	

Número	Nombre de la Unidad	Duración en Semanas	
2	Aprendizaje Supervisado	4	
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía	
Clasificación: <ol style="list-style-type: none"> <li>1. Metodologías clásicas de clasificación (árboles de decisión, SVM, naive bayes, KNN, otros).</li> <li>2. Evaluación del desempeño de clasificadores (bootstrap, crossvalidation).</li> <li>3. Comparación de modelos y clasificadores.</li> </ol>	El estudiante: <ul style="list-style-type: none"> <li>- Define qué es el aprendizaje supervisado.</li> <li>- Conoce diferentes metodologías de clasificación.</li> <li>- Aplica técnicas de clasificación sobre conjuntos de datos.</li> <li>- Interpreta los resultados obtenidos de diferentes clasificadores.</li> </ul>	[1] Cap. 4 Cap. 5	[3] Cap. 3.1-3.3, Cap. 4.1- 4.4, Cap. 4.6, Cap. 5, Cap. 6.1-6.5, 6.7

Número	Nombre de la Unidad	Duración en Semanas	
3	Aprendizaje No-Supervisado	4	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<ol style="list-style-type: none"> <li>1. Introducción y conceptos básicos de clustering.</li> <li>2. Algoritmos clásicos de clustering (k-means, clustering aglomerativo jerarquico, otros métodos).</li> <li>3. Elección de métodos de clustering.</li> <li>4. Evaluación de clusters.</li> </ol>		<p>El estudiante:</p> <ul style="list-style-type: none"> <li>- Define qué es el aprendizaje no-supervisado.</li> <li>- Conoce diferentes metodologías de clustering.</li> <li>- Aplica técnicas de clustering sobre conjuntos de datos.</li> <li>- Interpreta los resultados obtenidos de diferentes algoritmos de clustering.</li> </ul>	<p>[1] Cap. 8 Cap. 9</p> <p>[3] Cap. 3.9, Cap. 4.8, Cap. 6.6</p>

Número	Nombre de la Unidad	Duración en Semanas	
4	Análisis de Asociación	4	
Contenidos		Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<ol style="list-style-type: none"> <li>1. Generación de conjuntos de elementos frecuentes.</li> <li>2. Reglas de asociación.</li> </ol>		<p>El estudiante:</p> <ul style="list-style-type: none"> <li>- Define qué es análisis de asociación.</li> <li>- Conoce la minería de reglas de asociación y donde se aplica.</li> <li>- Aplica técnicas de asociación sobre conjuntos de datos.</li> <li>- Interpreta los resultados obtenidos de diferentes tipos de análisis de asociación.</li> </ul>	<p>[1] Cap. 6 Cap. 7</p> <p>[3] Cap 3.4-3.6, Cap. 4.5</p>

### Bibliografía

1. P. Tan, M. Steinbach, V. Kumar. Introduction to Data Mining. Addison-Wesley. 2006.
2. H. Samet. Foundations of Multidimensional and Metric Data Structures. Morgan Kaufmann. 2006.
3. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems) by Ian H. Witten and Eibe Frank .
4. P. Zezula, G. Amato, V. Dohnal, M. Batko. Similarity Search: The METric Approach. Springer. 2006.

Vigencia desde:	Otoño 2014
Elaborado por:	Bárbara Poblete