

## Syllabus TICS 411 Minería de Datos

Unidad académica	Pregrado		
Carrera o programa	Ingeniería Civil Informática		
Año	2020	Semestre	2
Profesor	Federico Jensen Sebastián Hughes Sebastián Moreno	Email	federico.jensen@edu.uai.cl sebastian.hughes@edu.uai.cl sebastian.moreno@uai.cl
		Horario de atención	Bajo demanda previa, coordinación por correo
Ayudante	Por decidir	Email	Por decidir
Créditos SCT-Chile	6	Total horas	180
Horas de Docencia Directa		Horas de Trabajo Autónomo	
Cátedra	Laboratorio	Ayudantía	
45	0	23	112
Tipo de Asignatura	Mayor.		
Línea curricular/Área	Informática		
Pre-requisitos	Base de Datos (TICS320)		
Descripción de la asignatura	<p>En la actualidad, las empresas colectan y administran una gran cantidad de datos. Pero, ¿con qué fin? En un principio, tener la mejor y más moderna infraestructura tecnológica, se veía como una forma de lograr una ventaja estratégica, sin embargo, luego se demostró que no era muy efectivo, ya que la infraestructura tecnológica hoy en día es un commodity que todas las empresas pueden y deben tener, por lo tanto, la infraestructura tecnológica (por ejemplo computadores), no es algo exclusivo de una empresa, y no genera una ventaja estratégica de negocio. El segundo paso entonces, fue la recolección de datos, ya que tener datos es tener información. Pero, nuevamente se observa que todas las empresas colectan datos, y por lo tanto, tampoco es exclusivo y no le confiere liderazgo en el mercado a una empresa.</p> <p>Hoy en día, la forma de procesar los datos está logrando ventajas competitivas y estratégicas en la toma de decisiones en cualquier organización. Esto impacta directamente tareas tales como el razonamiento bajo incertidumbre, identificación de patrones de comportamiento, predicción de fenómenos, identificación de factores más determinantes en un negocio, detección de tendencias, segmentación de grupos, etc.</p> <p>Para abordar las tareas previas, se requiere de nuevas técnicas para analizar inteligentemente datos generados en los negocios y convertirlos así en conocimiento valioso, actividad comúnmente conocida como “Minería de Datos”, cuyo componente principal se basa en modelos y técnicas de análisis de datos.</p>		
Competencias del egresado	(a) Aplicar conocimiento de matemáticas, ciencia e ingeniería. (b) Diseñar y conducir experimentos, como también para analizar e interpretar datos. (e) Identificar, formular y resolver problemas de ingeniería (j) Conocer temas contemporáneos (k) Usar técnicas, habilidades y herramientas modernas de ingeniería necesarias para la práctica ingenieril.		
Resultados de Aprendizaje	Al terminar el curso el alumno estará capacitado para: <ul style="list-style-type: none"> <li>- Identificar las etapas claves de la Minería de datos</li> <li>- Identificar y modelar las diferentes tareas de Minería de datos (ej, tareas predictivas, descriptivas, etc)</li> <li>- Analizar y explorar datos estructurados</li> <li>- Aplicar técnicas y modelos de análisis automático de datos, del tipo descriptivo y predictivo.</li> <li>- Evaluar y comparar diferentes técnicas y modelos.</li> </ul>		

Estrategias de enseñanza y aprendizaje
<p>El desarrollo del curso se llevará a cabo mediante las siguientes actividades:</p> <ul style="list-style-type: none"> <li>- <b>Clases expositivas:</b> realizadas por el profesor donde se revisarán los aspectos conceptuales, técnicas y modelos, y aspectos prácticos usando ejemplos en Python o R.</li> <li>- <b>Trabajos en clases:</b> ejercicios semanales con diferentes técnicas computacionales usando ejemplos en Python o R.</li> <li>- <b>Proyecto semestral:</b> realización de un proyecto semestral práctico (grupal) que resuelva algún problema de minería de datos utilizando diversas herramientas computacionales.</li> </ul>
Procedimientos de Evaluación de aprendizajes
<p>El curso considera un solo mecanismo de evaluación: un <i>proyecto grupal</i>. Este proyecto resolverá algún problema de minería de datos utilizando diversas herramientas computacionales, y se divide en tareas parciales durante el semestre.</p>
Unidades de la asignatura
<ol style="list-style-type: none"> <li>1. Introducción a la Minería de datos <ol style="list-style-type: none"> <li>1.1. Conceptos en análisis de datos</li> <li>1.2. El proceso de Minería de Datos</li> <li>1.3. Tipos de problemas: descriptivos, predictivos o reconocimiento de patrones.</li> </ol> </li> <li>2. Análisis Exploratorio de datos <ol style="list-style-type: none"> <li>2.1. Tipos y calidad de datos</li> <li>2.2. Pre-procesamiento de datos</li> <li>2.3. Medidas de similaridad y distancia.</li> <li>2.4. Outliers.</li> </ol> </li> <li>3. Modelamiento descriptivo <ol style="list-style-type: none"> <li>3.1. Introducción a los modelos descriptivos (aprendizaje no supervisado)</li> <li>3.2. Clustering particional, basado en densidad, jerárquico y difuso (K-means, DB-SCAN, aglomerativo, C-means)</li> <li>3.3. Evaluación de clusters</li> </ol> </li> <li>4. Modelamiento predictivo <ol style="list-style-type: none"> <li>4.1. Introducción a los modelos predictivos (aprendizaje supervisado)</li> <li>4.2. Clasificadores probabilísticos, algoritmos: Naive Bayes y Regresión Logística</li> <li>4.3. Clasificadores discriminativos, algoritmos: K-Nearest neighbors, árboles de decisión, SVM</li> <li>4.4. Evaluación de modelos predictivos: validación cruzada, Curvas ROC.</li> <li>4.5. Técnicas de ensamblados.</li> </ol> </li> </ol>
Evaluación
<p>La evaluación del curso involucra las siguientes actividades: Informe análisis exploratorio de datos (I1), Informe aprendizaje no supervisada (I2), e informe aprendizaje supervisado (I3):</p> <p>La nota final del curso se calcula como: <math>NF = 0.30 * I1 + 0.30 * I2 + 0.40 * I3</math></p> <p>Además, habrá un examen recuperativo, después del periodo de exámenes fijado por la Universidad, el cual puede cubrir toda la materia del semestre. Este examen recuperativo lo podrán dar cualquier alumno en forma voluntaria y cubrirá toda la materia de la asignatura. Por lo mismo, la no realización de un informe será evaluada con nota 1.0. En caso de dar este examen, la nota final del curso corresponderá a</p> $NF = 0.18 * I1 + 0.18 * I2 + 0.24 * I3 + 0.40 * ER$
Bibliografía
<p>Básica:</p> <ul style="list-style-type: none"> <li>• <b>Introduction to Data Mining.</b> Tan, Steinbach, and Kumar (2005).</li> <li>• <b>Data Science for Business.</b> Provost and Fawcett. (2013)</li> </ul> <p>Complementaria:</p> <ul style="list-style-type: none"> <li>• <b>Data Mining the textbook.</b> Charu Aggarwal. (2015)</li> <li>• <b>Pattern Recognition and Machine Learning.</b> Christopher Bishop. (2006)</li> </ul>
Notas
<p><b>La UAI está preocupada de mantener un ambiente de trabajo y aprendizaje inclusivos, y por lo tanto no se permitirá ninguna forma de discriminación. Cualquier tipo de discriminación por sexo, estado civil, embarazo, raza, apariencia física, capacidad diferente, orientación sexual, identidad de género, edad, nacionalidad, religión u otra característica legalmente protegida, está estrictamente prohibida.</b></p>