



UNIVERSIDAD CATÓLICA DEL NORTE

PROGRAMA DE CURSO

I. IDENTIFICACIÓN DEL CURSO	
Programa: Magister Ingeniería Informática	
Unidad responsable: Departamento de Ingeniería de Sistemas y Computación	
Nombre del curso: Data Mining	
Número de Créditos SCT: 10	
Código: Por definir	
Semestre en la malla: Por Definir	
Cursos Requisitos: Machine Learning	Requisito para: Web Intelligence
II. DESCRIPCIÓN DEL CURSO	
<p>El creciente interés en data mining está motivado por un problema común a varias disciplinas: como almacenar, acceder, modelar, y últimamente describir y entender conjuntos de datos masivos? Históricamente, diferentes aspectos de data mining han sido direccionados independientemente por varias disciplinas.</p> <p>Este curso introductorio de a data mining da una visión general del area y mezcla contribuciones desde la ciencia de la información, la ciencia de la computación, la estadística y los negocios. Bajo el marco de trabajo de un proceso de KDD y de una guía de desarrollo de proyectos de data mining, se estudian métodos de preparación de datos y técnicas avanzadas de minado de patrones en datos, que incluyen técnicas de clasificación, predicción, y agrupamiento. Los estudiantes del curso deberán aplicar tanto el proceso como las técnicas estudiadas a un problema original en el contexto de un mini-proyecto desarrollado durante el curso.</p>	
III. UNIDADES TEMÁTICAS	
Unidad 1: Introducción	
1.1 Que es data mining?	
1.2 Relaciones con otras áreas	
1.3 Aplicaciones exitosas	

- 1.4 Tipos de datos a ser minados
- 1.5 Tipos de patrones minados
- 1.6 Tareas primitivas de data mining
- 1.7 Integración Data Warehouse – Data Mining
- 1.8 Web Mining

Unidad 2: Proceso de Descubrimiento de Conocimiento en Bases de datos

- 2.1 El proceso KDD
- 2.2 La guía CRISP-DM
- 2.3 Entendimiento del Negocio
- 2.4 Entendimiento de los Datos
- 2.5 Preparación de los datos
- 2.6 Modelado de datos
- 2.7 Interpretación y evaluación de modelos
- 2.8 Implantación de los resultados

Unidad 3: Preparación de datos

- 3.1 Análisis de relaciones y dependencias entre variables
- 3.2 Muestreo de datos
- 3.3 Selección de atributos
- 3.4 Discretización de datos
- 3.5 Detección de datos anómalos
- 3.6 Limpieza, integración, transformación y reducción

Unidad 4: Algoritmos avanzados de data mining

- 4.1 Inducción de reglas de clasificación
- 4.2 Reglas de asociación. El algoritmo Apriori.
- 4.3 Máquinas de vectores de soporte

4.4 Regresión lineal y no lineal

4.4 Caso de estudio

Unidad 5: Clustering de datos

5.1 Que es el análisis de clustering?

5.2 Taxonomía de métodos de clustering

5.3 Clustering por particionamiento

5.4 Clustering jerárquico

5.5 Otras formas de clustering

5.6 Análisis de outliers

5.7 Caso de estudio

Unidad 6: Evaluación de lo aprendido en clasificación y predicción

6.1 Medidas de precisión y error

6.2 Evaluación de la precisión de un clasificador y predictor (cross validation, bootstrap)

6.3 Métodos de combinación (bagging, boosting)

6.4 Selección de modelos (estimación intervalos de confianza, curvas ROI)

Unidad 7: Herramientas de data mining

7.1 Excel

7.2 SPSS

7.3 WEKA

7.4 Clementine

7.5 Suite Pentaho

IV. RESULTADOS DE APRENDIZAJE

A. Comprensión de los conceptos básicos y aplicaciones principales de data mining tanto en el aspecto científico como de negocios.

B. Comprensión y utilización de técnicas avanzadas de data mining para resolver casos ficticios y reales.

C. Aplicación de una guía y métodos al desarrollo de un caso original.

D. Utilización de herramientas de software de análisis y preparación de datos.

V. RECURSOS BIBLIOGRÁFICOS

Textos Guías:

1. **J. Han & M. Kamber.** “Data Mining: Concepts and Techniques”, ISBN 13: 978-1-55860-901-3, 2006.
2. **P. Tan, M. Steinbach & V. Kumar.** “Introduction to Data Mining”, ISBN: 0321321367, 2006.

Textos o lecturas complementarias:

3. **H. Witten, E. Frank & M Hall.** “Data Mining: Practical Machine Learning Tools with Java Implementations”, Third Edition, ISBN: 978-0-12-374856-0 , 2011.
4. **D. Hand, H. Mannila & P. Smyth.** “Principles of Data Mining”, ISBN: 0-262-08290-X, 2001.

Artículos de Revistas