Universidad de Los Andes
Facultad de Ciencias Económicas y Sociales
Maestría en Estadística
Mérida – Venezuela

Materia: Introducción a la Minería de Datos (Electiva)	**Número de Unidades Crédito**: 02
Trimestre: V Trimestre	**Horas**: 32
Mención: Maestría en Estadística	

OBJETIVO GENERAL

Desarrollar destrezas básicas en la extracción de conocimiento a partir de grandes volúmenes de información.

OBJETIVOS ESPECÍFICOS

- Introducir al estudiante a los conceptos básicos de la minería de datos. Características particulares y relaciones con otras disciplinas.
- Revisar el tema de los Almacenes de Datos o Data Warehouse: diseño y manipulación. Tendencias actuales en manejo de persistencia de información.
- Comprender el proceso general de la minería de datos: preparación de los datos, aplicación de las técnicas de análisis y generación de reportes.
- Revisar las técnicas fundamentales de la Minería de Datos: Reglas de Asociación, Clasificación, Clustering, Predicción, Redes Neuronales, etc.
- Estudiar distintos campos de aplicación: Minería de Series de Tiempo, Minería de Web y de Textos, Detección de Fraudes…

PROGRAMA

1. Introducción:
   Conceptos básicos de minería de datos y aprendizaje automático. Campos de aplicación. Representación del conocimiento.
2. Almacenes de Datos:
   El problema de la persistencia de la información, Bases de datos SQL y noSQL. El almacén de datos como solución.
3. El proceso de la minería de datos:
   Preparación de los datos, aplicación de las técnicas de análisis y generación de reportes.
4. Algoritmos básicos:
   Análisis exploratorio de datos. Modelos estadísticos. Construcción de árboles de decisión y reglas de decisión.
5. Clasificación:
   Modelos de clasificación. Evaluación. Árboles de clasificación. Métodos bayesianos. Regresión logística. Redes Neuronales. Support Vector Machines.
6. Clustering (agrupamientos):
   Métodos de clustering y medidas de afinidad. Métodos de particionamiento. Métodos jerárquicos.
7. Reglas de asociación:
   Motivación y estructura de las reglas de asociación. Reglas de asociación unidimensionales. Reglas de asociación generales.
8. Minería de textos y web:

Extracción de datos de la Web. Minería del uso. Minería del contenido. Minería de la estructura.

ESTRATEGIAS METODOLÓGICAS

- El curso tendrá un enfoque completamente práctico en el que se ilustrarán y revisarán los conceptos fundamentales de aprendizaje automático a partir de grandes cantidades de información.
- La aproximación fundamental consistirá en la revisión de ejemplos de aplicaciones de las distintas técnicas a considerar, y el posterior análisis y discusión de los resultados. Se hará uso exclusivo de herramientas de software libre basadas en los lenguajes de programación, tales como R y Python, que permiten la mayor flexibilidad posible en la extracción, manipulación y almacenamiento de los datos.

EVALUACION

La evaluación consistirá en el desarrollo de proyectos que permitan integrar y aplicar los contenidos. Este proyecto se realizará por etapas que serán revisadas cada semana:

- Realización de dos guías de ejercicios (60%)
- Mini-proyecto final (40%).

BIBLIOGRAFIA

- Applied Data Mining for Business and Industry Second Edition . Paolo Giudici y Silvia Figini. John Wiley & Sons, 2009.
- Business Intelligence: Data Mining and Optimization for Decision Making . Carlo Vercellis. John Wiley & Sons, 2009.
- Data Mining : Practical Machine Learning Tools and Techniques Third Edition. Ian H. Witten . Eibe Frank y Mark A. Hall . Morgan Kauffman – Elsevier, 2011.
- Introduction to Machine Learning Second Edition . Ethem Alpaydın . The MIT Press, 2010.
- R in a Nutshell . Joseph Adler . O'Reilly Media, 2010.

Enlaces básicos

- Machine Learning Open Source Software http://mloss.org
- Journal of Machine Learning Research http://jmlr.csail.mit.edu/
- El Blog del Geomarketing y el Data Mining http://geomarketingspain.blogspot.com/
- Some Datasets Available on the Web http://www.datawrangling.com/some-datasets-available-on-the-web
- Data Mining Community's Top Resource http://www.kdnuggets.com

Mayo 2011
FPR/Lucina