

**UNIVERSIDAD DE PUERTO RICO
RECINTO UNIVERSITARIO DE MAYAGUEZ
DEPARTAMENTO DE MATEMATICAS**

**COMP6838: MINERIA DE DATOS (DATA MINING)
(COMP 6315)**

Rev. Septiembre 15 del 2010

Instructor: [Edgar Acuna](#)

Objetivos del curso: Entender los conceptos fundamentales para llevar a cabo minería de datos y descubrimiento de conocimiento en base de datos. Experimentar algunos algoritmos más usados en minería de datos en conjuntos de datos reales.

Horario del curso: M y J de 2 -3.15pm en M118.

Pre-requisitos del curso: Haber tomado dos cursos donde se hayan visto conceptos estadísticos. Tener algún conocimiento de matrices, sistemas de bases de datos y de algún programa de computación.

Lenguaje de programación: Programas gratuitos de minería de datos tales como R, Weka, RapidMiner(Yale) y Orange. Microsoft SQL (Analysis Services) también puede ser usado.

Oficina: M314.

Horas de oficina: M 7.30 a 8.30am y J de 9.30 a 10.30am L de 7.30 a 9 am y de 11 a 1.00pm.

Extension: x3287

Correo electronico del Profesor: eacuna@uprm.edu ,
edgar@math.uprm.edu, edgar@cs.uprm.edu eacunaf@yahoo.com

Pagina web del curso: math.uprm.edu/~edgar/comp6838.html

Ayudante: Roxana Aparicio (M309).

Contenido del curso

I. Introducción (3 horas)

Que es minería de datos?

Tipos de datos.

Aplicaciones de minería de datos.
Clasificación de sistemas de minería de datos.
Problemas principales en minería de datos.

II. Pre-procesamiento de datos (15 horas)

Limpieza de datos.
Tratamiento de datos faltantes
Integración de datos y Transformación
Reducción de la dimensionalidad : Selección de características, Extracción de Características.
Discretización
Selección de casos y detección de ruido.

III. Visualización (5 horas)

Plot de coordenadas paralelas
Surveyplots
Visualización radial
Visualización starcoordinates

VI. Detección de “Outliers” (5 horas)

Métodos estadísticos de detección de “outliers”
Métodos para detectar “outliers” basados en distancia.
Métodos para detectar “outliers” basados en densidad local

EXAMEN PARCIAL

IV Clasificación Supervisada (9 horas)

Clasificación Bayesiana
Precisión de un clasificador
Clasificación usando “naïve Bayes”
Clasificación usando análisis discriminante Lineal
Clasificación usando arboles de decisión
Clasificación usando vecinos cercanos (k-nn)
Clasificación usando regresión logística
Clasificación usando redes neurales

V. Analisis por conglomerados (7 horas)

Métodos de Particionamiento: El Algoritmo de k-medias y el PAM
Métodos Jerárquicos: Algorerativos y divisivos
Mapas auto-organizantes (SOM) de Kohonen
Metodos de conglomerados basados en modelos
Validación de conglomerados

Texto:

Tan, P.N., Steinbach, M. y Kumar, V. (2006) Introduction to data Data Mining. Pearson Addison Wesley, Boston.

Han, J. y Kamber, M. (2006). Data Mining: Concepts and Techniques. Second Edition. Morgan Kaufman Publishers. San Francisco. La primera edición de este libro esta disponible gratis en la internet en <http://www.cs.wmich.edu/~yang/teach/cs595/han>

References

1. Hastie, T, Tibshirani, R & Friedman, J. (2001). *Elements of Statistical Learning*. Springer-Verlag.
2. Kantardzic M. (2003). *Data Mining: Concepts, Models, methods, and Algorithms*. John Wiley. New York.
3. Williams, G. [Data Mining Desktop Survival Guide](#), on-line book (PDF).
4. Witten, I.H., and Frank, E. (2005), *Data Mining: Practical Machine Learning tools and techniques*. Morgan Kaufman, San Francisco.

Evaluación

1	Tareas (4)	40%
2.	Examen Parcial	30%
3.	Proyecto incluyendo presentación oral	30%

LINKS

<http://www.cs.sfu.ca/~han/dmbook> Slides del libro de Han

<http://www.eas.asu.edu/~mining03>. Curso de Data Mining por el professor Liu.

<http://www.liacc.up.pt/~ltorgo/DataMiningWithR/>. Libro de Data Mining con R por Luis Torgo.

<http://www.kdnuggets.com/dmcourse> Curso en Data mining por kdnuggets

<http://www.galaxy.gmu.edu/stats/syllabi/DMLIST.html>. Varios links a Data Mining

http://www.pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_1.html . Notas de Data Mining de la Queen's University de Belfast.

<http://www.scd.ucar.edu/hps/GROUPS/dm/dm.html> Varios links a Data Mining

<http://ocw.mit.edu/OcwWeb/Sloan-School-of-Management/15-062Data-MiningSpring2003/LectureNotes/>. Curso de Data Mining en MIT

<http://www.cise.ufl.edu/class/cis6930fa03dm/notes.html> Curso de Data Mining, CISE, University of Florida

<http://www.cs.uu.nl/docs/vakken/adm/>. Notas en data Mining del Profesor Felders

<http://www.cs.unc.edu/Courses/comp290-90-f04/> Notas en Data Mining del Profesor Wang

<http://www.cs.pdx.edu/~timm/dm/>. Notas en Data Mining de Pórtland University

<http://www.stats.ox.ac.uk/~vos/DataMining.html> Data Mining for Bioinformatics with R

<http://ljsavage.wharton.upenn.edu/~stine/mich/>. Notas de Data Mining de University of Pennsylvania.

<http://www.galaxy.gmu.edu/stats/syllabi/INFT979.spring2001.html> Notas en data mining del professor Wegman.