

	<p>SEMINARIO: Métodos Descriptivos en Minería de Datos Tutor: Dr. Oldemar Rodríguez Rojas</p> <p>Correo electrónico: <a href="mailto:correo@oldemarrodriguez.com">correo@oldemarrodriguez.com</a> Documentos del seminario en: <a href="http://www.oldemarrodriguez.com">http://www.oldemarrodriguez.com</a></p>	<p>Naturaleza: Teórico/Práctico Modalidad: Seminario Tipo de Seminario: Regular Horas totales: 16 Horas docencia directa: 14 Requisito: El/la estudiante debe asistir con su computador portátil.</p>
---	--	---

## **PROGRAMA DEL SEMINARIO**

### DESCRIPCIÓN DEL SEMINARIO

En este seminario se presentarán los principales conceptos y métodos en Minería de Datos. El énfasis principal del seminario será examinar dichos métodos desde un punto geométrico y de sus aplicaciones concretas. Se le dará especial importancia al uso de los conceptos de minería de datos en aplicaciones reales con bases de datos de gran tamaño, para esto se utilizarán los programas especializados en Minería de Datos como *FactoMineR* sobre la plataforma de software libre *R* y *RComander*.

### OBJETIVOS

En este seminario el estudiante:

1. Entenderá la necesidad de la utilización de modelos, algoritmos, software especial para el descubrimiento de conocimiento en grandes volúmenes de datos.
2. Conocerá la Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM.
3. Conocerá la metodología del ciclo de desarrollo usado para el descubrimiento del conocimiento en grandes bases de datos (KDD – “Knowledge Discovery in Databases”).
4. Entenderá las diferencias entre: estadística, análisis de datos, recuperación de la información, ML – “Machine Learning” y minería de datos.
5. Conocerá los principales modelos, técnicas y algoritmos utilizados para descubrir el conocimiento en grandes volúmenes de datos.
6. Utilizará el *FactoMineR* y *RComander* sobre la plataforma *R* para analizar ejemplos con datos reales.

### CONTENIDO

1. Conceptos de la Minería de Datos
  - a. Definiciones básicas
  - b. Diseño de bodegas y mercados de datos
  - c. Herramientas OLAP
  - d. Metodología para el Desarrollo de Proyectos en Minería de Datos *CRISP-DM*.
2. Análisis Exploratorio de Datos
  - a. Tipos de variables
  - b. Estadísticas básicas y matriz de correlaciones
  - c. Tablas de datos y datos atípicos
  - d. Regresión lineal
  - e. Aplicaciones en casos reales con *RComander* sobre la plataforma *R*
3. Métodos de condensación de la información
  - a. Análisis en Componentes Principales – ACP (PCA, Karhunen-Loeve o K-L Method)
    - Plano principal
    - Círculo de correlaciones
    - Dualidad y sobre-posición de gráficos
    - Análisis Factorial de Correspondencias Múltiples
  - b. Aplicaciones en casos reales con *FactoMineR*

4. Clasificación Jerárquica Aglomerativa
  - a. ¿Qué es “cluster analysis”?
  - b. Clasificación Jerárquica Aglomerativa
    - Distancias y matrices de distancias
    - Agregaciones
    - Jerarquías binarias
    - Jerarquías Binarias sobre las Componentes Principales
  - c. Aplicaciones en casos reales con *FactoMineR*

## METODOLOGÍA

- Todo el material del seminario será presentado por el tutor.
- Se realizarán prácticas/laboratorio de minería de datos con ejemplo concretos.

## FECHAS IMPORTANTES

- El Seminario se realizará los días 26 y 27 del julio del 2012 de 8am a 5pm en el Hotel Crowne Plaza Corobicí, Sala Cahuita 2.

## BIBLIOGRAFÍA

1. Berry M. and Linoff G. “Data Mining Techniques”. John Wiley & Sonsa, 1997.
2. Bry X. “Analyses factorielles simples”, Ed. Economica, Paris, 1995.
3. Hastie, Tibshirani and Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, 2009.
4. John M. Chambers. Programming with R: Software for Data Analysis. Springer, Stanford University, Palo Alto, 2008.
5. Giudici Paolo. “Applied Data Mining: Statistical Methods for Business and Industry”. Wiley, 2005.
6. Graham Williams, Data Mining with Rattle and R. Springer, New York, 2011.
7. Dunhan M. “Data mining: Introductory and Advanced Topics”. Prentice Hall, 2002.
8. Han J. and Kamber M. “Data Mining: concepts and techniques”, Morgan Kaufman Publishers 2001.
9. Jambu M. “Introduction au Data Mining: Analyse Intelligente des données”. Eyrolles, Paris, 1999.
10. Mirkin Boris. “Clustering for Data Mining, a data recovery approach”. Chapman & Hall. Boca Raton FL, 2005.
11. Owen Jones, Robert Maillardet and Andrew Robinson. Introduction to Scientific Programming and Simulation using R. Chapman & Hall/CRC Taylor & Francis Group, FL. 2009.
12. R Development Core Team. “R: A Programming Environment for Data Analysis and Graphics”. The R Project for Statistical Computing, 2010. <http://www.r-project.org/>
13. R Development Core Team. “Writing R Extensions”. The R Project for Statistical Computing, 2010. <http://www.r-project.org/>
14. Rodríguez O, P.J.F. Groenen S. Winsberg and E. Diday. "I-Scal: Symbolic Multidimensional Scaling of Interval Dissimilarities". COMPUTATIONAL STATISTICS & DATA ANALYSIS the Official Journal of the International Association for Statistical Computing, London, 2006.
15. W. J. Braun and D. J. Murdoch, A First Course in Statistical Programming with R. The University of Western Ontario, 2007.

Repositorio de Bases de Datos de ejemplos: <http://www.ics.uci.edu/~mlearn/MLRepository.html>