



PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de
Formación en Minería de Datos

**Métodos Predictivos
en Minería de Datos**



(506) 2268.8823 - (506) 8708.9091



info@promidat.com



facebook.com/oldemarrodriguez



www.promidat.com

Tutor: El curso será impartido por Dr. Oldemar Rodríguez graduado de la Universidad de París IX y con un postdoctorado en Minería de Datos de la Universidad de Stanford.

Duración: Cuatro semanas.

Descripción:



En este curso se presentarán los principales métodos en Minería de Datos, especialmente enfocados en métodos predictivos, conocidos también como métodos de aprendizaje supervisado. El énfasis principal del curso será examinar dichos métodos desde un punto de vista algorítmico y de sus aplicaciones en casos reales. 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 son la plataforma de desarrollo R y el paquete **predictoR**.

¿A quién va dirigido?

Se orienta especialmente a analistas de riesgo, marketing y mercadeo, personal involucrado en proyectos de pronóstico y predicción o en proyectos de “Business Intelligence”, administradores de bases de datos, analistas de sistemas, profesionales en estadística o economía y en general a personas que, basados en los datos de la organización, colaboran en la toma de decisiones.

Objetivos:

En este curso el estudiante será capaz de:

1. Comprender la diferencia entre modelos de aprendizaje supervisado (minería predictiva) y modelos de aprendizaje no supervisado (minería descriptiva).
2. Comprender la diferencia entre bases de datos de aprendizaje y bases de datos de “testing”.
3. Comprender la necesidad de la utilización de modelos, algoritmos, software para predecir el comportamiento futuro.
4. Conocer los principales modelos predictivos, técnicas y algoritmos utilizados para predecir conductas a partir de grandes volúmenes de datos históricos.

5. Utilizar la plataforma *R* y el paquete **predictoR** para analizar y desarrollar ejemplos con datos reales.

Metodología:

Basado en la teoría y en la aplicación directa de los conceptos aprendidos. Para esto se dispondrán de las siguientes herramientas.

- Una vídeo conferencia semanal, las cuales quedarán grabadas en Webex, para que los alumnos la puedan acceder en cualquier momento.
- Trabajos prácticos semanales.
- Foros para plantear dudas al tutor y compañeros.
- Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Desarrollar proyectos de Minería de Datos que involucren predicción utilizando modelos predictivos.

Contenido:

1. Conceptos de la Minería de Datos Predictiva

- a. Conceptos y diferencias entre aprendizaje supervisado y aprendizaje no supervisado
- b. Diseño de bases de datos de aprendizaje
- c. Diseño de bases de datos de testing
- d. Variables cuantitativas y variables cualitativas
- e. ¿Cómo evaluar la calidad de un modelo predictivos?
- f. Cálculo de la Matriz de confusión e índices de calidad
- g. Curvas ROC
- h. Aplicación con datos reales con **predictoR**

2. Método de los K vecinos más cercanos

- a. Estructura General del método
- b. El mejor valor de K

- c. Algoritmo de Aprendizaje
- d. Aplicación con datos reales con **predictoR**

3. Máquinas Vectoriales de Soporte

- a. Hiperplano de separación de las clases
- b. Vectores de soporte
- c. Función discriminante lineal
- d. ¿Cómo resolver un Problema Optimización?
- e. MVS no linealmente separables
- f. Núcleos en Máquinas Vectoriales de Soporte
- g. Aplicación con datos reales con **predictoR**

4. Árboles de Decisión (Método CART)

- a. Algoritmos ID3, C4.5, C5.0 y CART
- b. Árboles de auto-regresión
- c. Aplicación con datos reales con **predictoR**

5. Métodos de consenso y de Potenciación

- a. Métodos de Consenso (Bagging)
- b. Bosques Aleatorios (Random forests)
- c. Métodos de impulso (Boosting)
- d. Métodos de Potenciación (ADA Boosting)
- e. Aplicación con datos reales con **predictoR**

Bibliografía:

1. Berry M. and Linoff G. "Data Mining Techniques". John Wiley & Sons, 1997.
2. Hastie, Tibshirani and Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, 2009.
3. Giudici Paolo. "Applied Data Mining: Statistical Methods for Business and Industry". Wiley, 2005.
4. Dunham M. "Data mining: Introductory and Advanced Topics". Prentice Hall, 2002.

5. Han J. and Kamber M. "Data Mining: concepts and techniques", Morgan Kaufman Publishers 2001.
6. Mirkin Boris. "Clustering for Data Mining, a data recovery approach". Chapman & Hall. Boca Raton FL, 2005.
7. 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/>
8. R Development Core Team. "Writing R Extensions". The R Project for Statistical Computing, 2010. <http://www.r-project.org/>