

## **ANEXO I. Data Mining.**

### **TEMARIO:**

#### 1. Introducción

- Objetivos del curso
- Motivación
- Ejemplos
- Definición
- Parámetros del crecimiento de datos
- Definición

#### 2. Mejoramiento de la calidad de datos

- Orígenes
- Clasificación
- Metodología PLEDEVSA
- Métricas de calidad de datos
- Métodos de limpieza
- Estándares y normas ISO
- Herramientas comerciales y software libre
- Desarrollo de un Plan de Trabajo
- Ejercicio con Excel

#### 3. Bases de datos

- Características de una base de datos para Minería de Datos
  - La clase
  - El tamaño
  - El valor comercial
- Tipos de variables en una base de datos
- Cáncer en las bases de datos
  - Maldición de la dimensionalidad
  - Desbalance de clases

#### 4. Modelos

- Definición
  - Conjunto de entrenamiento

- Conjunto de prueba
- Cáncer del Modelo
- Evaluación del modelos
  - El error en el conjunto de prueba
  - Métodos de cálculo de error
- Matriz de confusión

#### 5. Técnicas de análisis avanzado

- CRISP-DM
  - Historia
  - Etapas generales
  - Fases específicas
  - Detalles en los documentos de entrega de proyecto basados en CRISP-DM
- SAS SEMMA
  - Comparación contra CRISP-DM.
- Árboles de decisión
  - Entropía
  - CHAID / CART
- Clustering
  - Segmentation vs Clustering vs Grouping
  - K-Medias
  - Vecinos más cercanos
- Reglas de asociación

#### 6. Proyecto de Data Mining

- Plan de trabajo de un proyecto de Data Mining
- Estimación de tiempos
- Cálculos de recursos

#### 7. Ejemplos de Minería de Datos.

- Detección de la clase
- Clasificación de las variables
- Ejercicios usando Weka
- Proyecto de Minería de Datos.
- Interpretación.
- Explicación de un proyecto de inicio a fin
- Como vender un proyecto de Minería de Datos.
- Selección del cliente.
- Valor agregado en Data Mining.

#### 8. Ejercicio práctico en software especializado

- Introducción al software.
- Características del producto.
- Arquitectura de configuración.
- Desarrollo e Implementación de un proyecto en la herramienta.
- Uso de técnicas de Minería de Datos con software especializado.
- Manejo de cubos de información.

#### 9. Introducción a Big Data

- Cuando iniciar un proyecto de Big Data
- Organigrama en una estructura para la implementación de un proyecto .
- Software usado

#### 10. Técnicas de graficación.

- Teoría de colores.
- Metodologías de graficación
- Ejercicio de graficación.

## **Instructor**

Dr. Luis Carlos Molina Félix

Sus estudios de maestría los realizó en el Instituto de Matemáticas y Computación de la Universidad de São Paulo (Brasil) y de doctorado en el Departamento de Lenguajes y Sistemas Informáticos de la Universidad Politécnica de Cataluña (España).

De 1999 al 2006 impartió el curso a distancia de Data Mining en la Universitat Oberta de Catalunya (UOC) en España. Es coautor junto con Ramón Sangüesa del libro: "Data Mining: Una Introducción" editado por la Fundación de la UOC. También es responsable del módulo de Data Mining en el Máster presencial de Ingeniería de Software de la Fundación Politécnica de Cataluña.

Ha trabajado en España, Brasil, Portugal como consultor para varios proyectos de Data Mining entre lo que se destacan de Telefónica España, del banco La Caixa, de la Universitat Oberta de Catalunya y de la consultora Accenture. En México se ha desempeñado como investigador y consultor en varios proyectos tales como Instituto Mexicano del Petróleo, BBVA Bancomer, Banejercito, Banorte, Consejo Coordinador Estatal de Seguridad Pública, Bancomer, Banorte, Iusacell, Oxxo, entre otros.

Cuenta con más de 20 publicaciones arbitradas de nivel internacional. Ha sido investigador huésped por la Universidad Atrás dos Montes, en Portugal y ha dado más de 50 conferencias sobre el tema de Data Mining. Desde el 2005 es director de Corporación Power Builders SC.