

# MINERIA DE DATOS EN LA PRACTICA: 2018 II

**Maestro:** Chris Stephens

**Oficina:** Centro de Ciencias de la Complejidad (frente a Universum), primer piso, aulas de enseñanza.

**Teléfono:** Celular – 5538151775.

**E-Mail:** stephens@nucleares.unam.mx

**Asistentes:** Jonathan Easton, Hugo Flores Huerta, David Ruíz

**Horario:** 4pm-7pm lunes

**Objetivo:** Aprender la minería de datos a través de trabajar con problemas del mundo real usando datos reales de varios sectores, como salud, educación, negocios, criminalidad etc. El propósito principal es aprender como modelar datos - es decir ser científico de datos - más que aprender un compendio de diferentes algoritmos. Se asignará a cada estudiante un proyecto que se espera poder convertir a una publicación.

## Contenido:

**Semanas 1 - 3:** Qué es la minería de datos? Qué son las metas?; “Big” data versus “Deep” data; Ciencia de los datos; Predicción versus perfilación; Preguntas en la minería de datos: Quien?, Qué? Cuando? Dónde? Porque?; Predicción y los sistemas complejos; El “Paisaje de Predictabilidad”; Regresión versus clasificación; Modelación; Medidas del éxito; Pasos en la minería de datos - Integridad de los datos, Selección de características, Algoritmos de Predicción - Naive Bayes.

**Semanas 4-8: Ejemplos reales** - Salud pública - predicción de gastos médicos mayores; predicción de factores de riesgo de obesidad y de diabetes; estilos de vida; Predicción de mortalidad después de un derrame y sus factores de riesgo. Negocios - predicción de compradores de polizas de seguro; quienes son tus mejores clientes? Adquisición, retención, ventas adicionales. Educación: Predicción del número de estudiantes de nuevo ingreso en una universidad; Predicción de la eficiencia terminal en las prepas; predicción de patrones de criminalidad; seguridad en redes. Redes sociales - procesamiento de lenguaje natural.

**Semanas 9-12:** Minería de datos espaciales; Inferencia a través de co-ocurrencias; Redes complejas inferenciales. Ecología: Predicción de los reservorios de una enfermedad emergente; Minería dinámica de datos; Optimización de intervenciones. Otras técnicas - Árboles de decisión; Reglas de asociación; Redes neuronales; Deep learning; Support vector machines; Series de tiempo; Hidden Markov models.