

INTRODUCCIÓN A LA MINERÍA DE DATOS

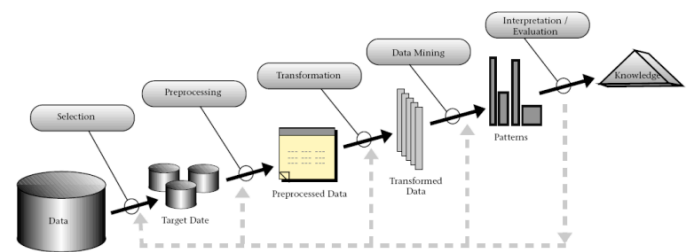
Bárbara Poblete
2011

HOY

- Motivación
- Un poco sobre la Metodología del Curso
- Planificación Tentativa

¿QUÉ ES LA MINERÍA DE DATOS?

- Descubrir automáticamente información útil en grandes repositorios de datos



KNOWLEDGE DISCOVERY IN DATABASES

¿POR QUÉ HACER MINERÍA DE DATOS?

- Punto de vista comercial y
- Punto de vista científico



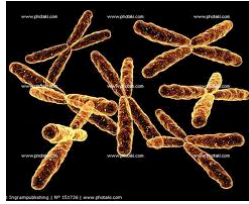
¿POR QUÉ HACER MINERÍA DE DATOS?

- ¿Motivación Comercial?
- Recolección de MUCHOS datos comerciales:
- Datos Web, e-commerce
- Compras en tiendas
- Transacciones en Bancos/ Tarjetas de Crédito



¿POR QUÉ HACER MINERÍA DE DATOS?

- ¿Motivación Científica?
- Datos (observaciones) recolectadas a gran velocidad (GB/hr)
- Telescopios, Satélites, Requerimientos Web, ADN, etc ([Google Flu Trends](#))

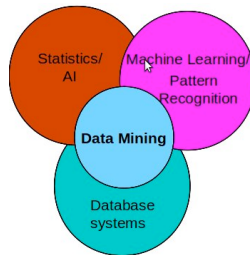


¿QUÉ (NO) ES MINERÍA DE DATOS (DM)?

- ¿Qué no es DM?
- ¿Qué sí es DM?

ORÍGENES DE LA DM

- Une ideas de ML/AI, reconocimiento de patrones, estadística y BD
- Enfoques tradicionales fallan con datos masivos



MÉTODOS DE DM

- Predictivos
- Descriptivos

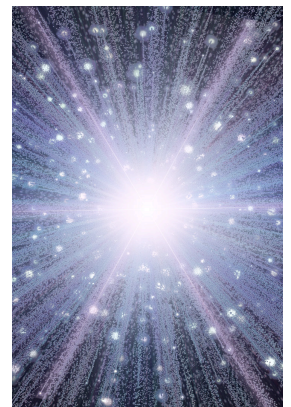


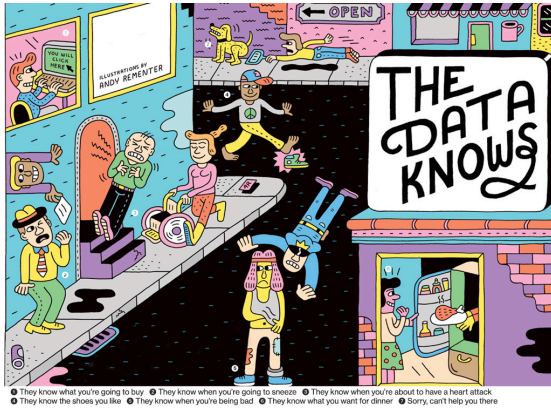
MÉTODOS DE DM

- Clasificación (Predictivo)
- Clustering (Descriptivo)
- Descubrimiento de Reglas de Asociación (Descriptivo)
- Descubrimiento de Patrones Secuenciales (Descriptivo)
- Regresión (Predictivo)
- Detección de Desviación (Predictivo)

BIG BANG

- 2006 Hadoop
- Análisis de datos masivos al alcance de todos (cientos de start-ups)





THE DATA KNOWS

Noticia en Businessweek

OBJETIVOS DEL CURSO

- Entender técnicas de DM que se aplican en el mundo real
- Conocer el estado del arte actual
- Auto-aprendizaje e investigación (usando múltiples fuentes)
- Saber aplicar técnicas de DM a problemas nuevos

METODOLOGÍA DEL CURSO

- Nota por Proyecto (2 Presentaciones)
- Nota Trabajo personal
 - 1 Mini-presentación (10 min)
 - Contribución de material (bloggeo)
 - Asistencia Tutorial WEKA
- 2 Controles