

# Ingeniería del conocimiento

---

## **Sesión 1**

### **¿Por qué estudiar aprendizaje automático?**

# Agenda

---

- ¿Qué vamos a ver en la asignatura?
- ¿Para qué sirve todo esto?
- ¿Cómo aprobar la asignatura?

# Extracción del conocimiento

---

# Extracción del conocimiento

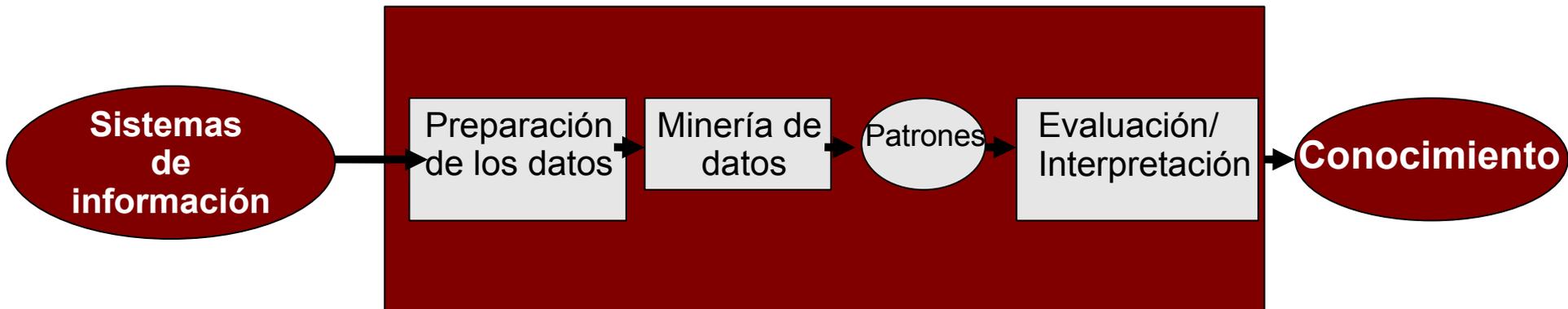
---

- Concepto de minería de datos o data mining:
  - *Proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos - Witten & Frank*
- Nuestro objetivo es pues convertir datos en conocimiento.
- Otras formas de llamarlo:
  - Intelligent Data Analysis
  - KDD – Knowledge Discovery in Databases

# KDD

---

- Normalmente nos referimos a KDD como todo el proceso de extracción del conocimiento, donde la minería de datos sería una de las fases.



# Datos

---

- ¿Cómo nos llegan esos datos?
  - Datos estructurados, semiestructurados o no estructurados.
- Bases de datos relacionales, html, xml, texto plano...
- O incluso podemos tener datos multimedia

# Modelos

---

- Nuestro objetivo es analizar los datos para extraer conocimiento.
- Este conocimiento puede ser en forma de reglas, relaciones, asociaciones, patrones o modelos.

# Modelos

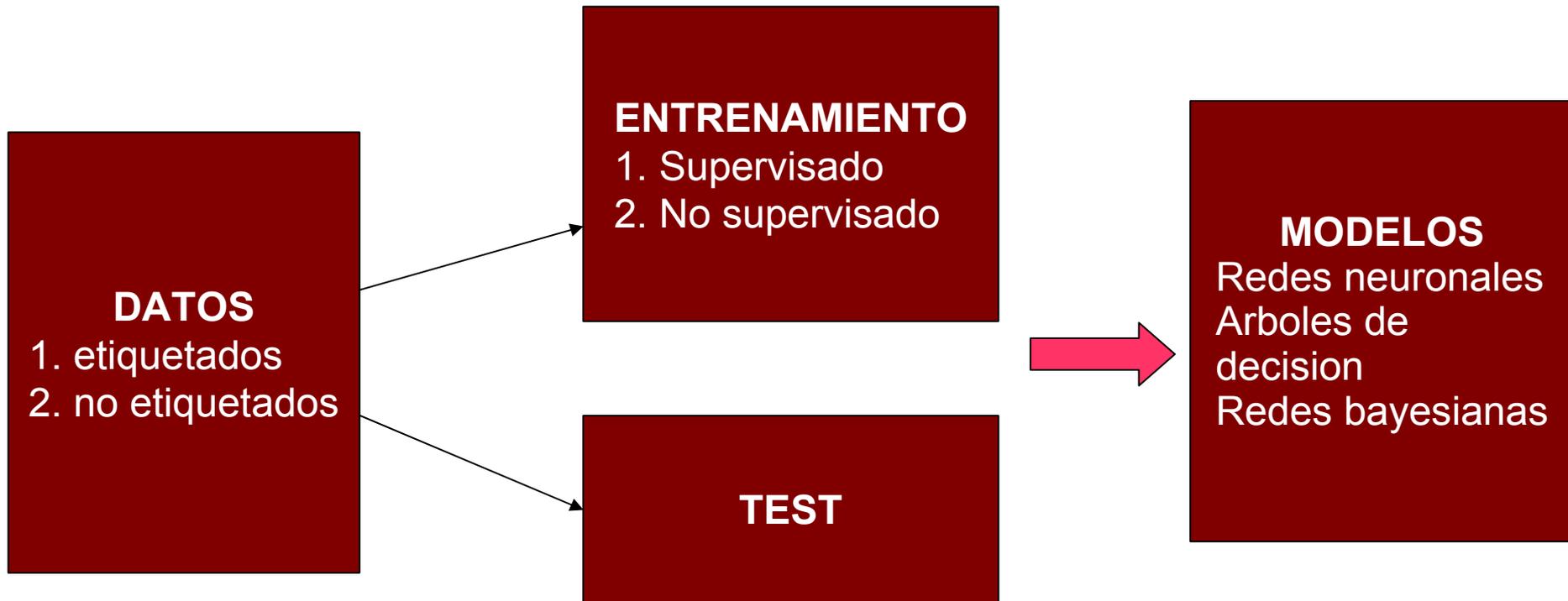
---

- Los modelos pueden ser predictivos o descriptivos
- Modelos predictivos: corresponden a tareas de clasificación y regresión
  - Árboles de decisión (o de clasificación)
  - Naive Bayes
  - Redes neuronales
  - ...
- Modelos descriptivos: agrupamiento (clustering), reglas de asociación, ...

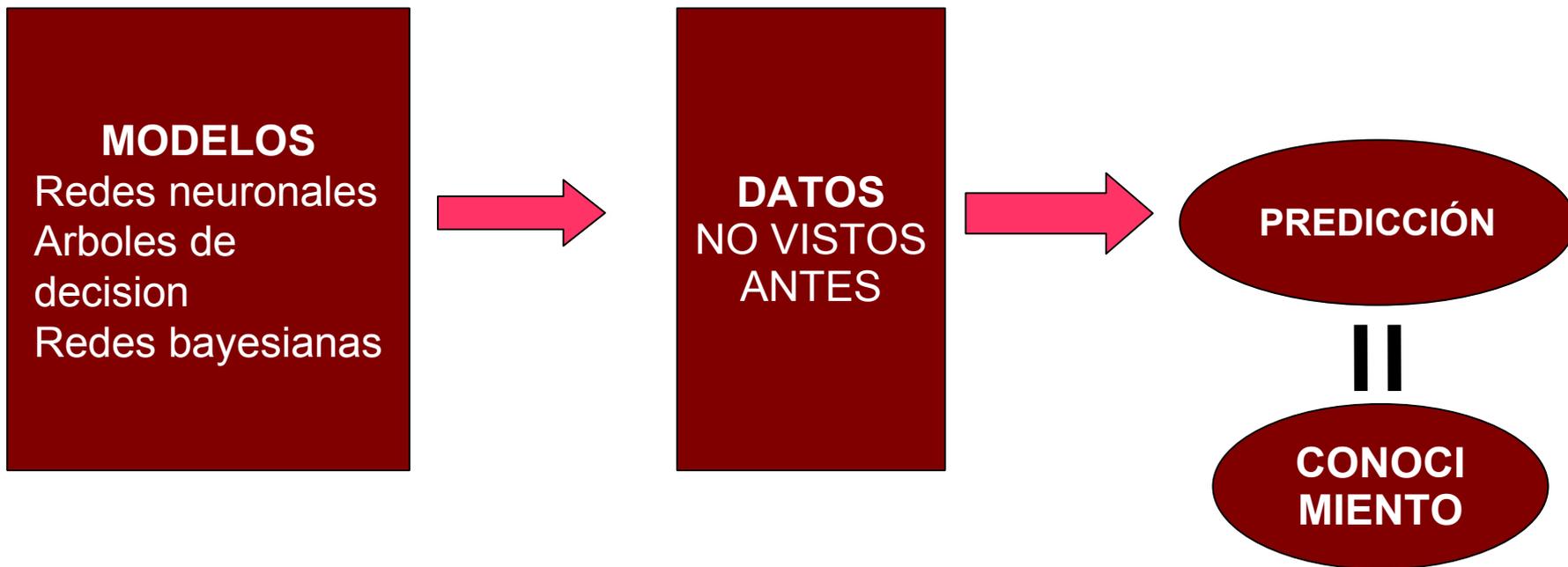
# ESQUEMA GENERAL DEL PROCESO

---

# ESQUEMA GENERAL DEL PROCESO



# ESQUEMA GENERAL DEL PROCESO



# Y ya se sabe..

---

**CONOCIMIENTO = €**

# ¿Para qué sirve todo esto?

## Aplicaciones

---

# EXTRACCIÓN DE CONOCIMIENTO DE GRANDES BASES DE DATOS

---

- **DATA WAREHOUSE** - Es una técnica para consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones (bases de datos de empleados de una empresa)
- **DATA MINING** - En este sentido un sistema de minería de datos es una tecnología de soporte para usuario final, cuyo objetivo es extraer conocimiento útil y utilizable a partir de la información contenida en las bases de datos de las empresas.

# EXTRACCIÓN DE CONOCIMIENTO DE GRANDES BASES DE DATOS

---

- Los objetivos:
  - analizar factores de influencia en determinados procesos,
  - predecir variables o comportamientos futuros
  - segmentar o agrupar items similares,
  - además de obtener secuencias de eventos que provocan comportamientos específicos.

# NEGOCIOS

---

- **CRM** - la estrategia de negocio centrada en anticipar, conocer y satisfacer las necesidades y los deseos presentes y previsibles de los clientes.
- Predicción automatizada de tendencias y comportamientos – Marketing predictivo y personalizado.
- Nos regalan tarjetas en todos los sitios.

# CIENCIA

---

- **Astrofísica** - Ingentes cantidades de datos – Clasificación, predicción. Proyecto telescopio MAGIC
- **Diagnóstico en Psiquiatría y Medicina:** Buscar en qué se parecen los pacientes de alguna enfermedad, es decir, aprender por casos que ya han sido diagnosticados. Proyecto del DSM IV
- **Predicción en psicología** - proyecto de maltratadores.
- **Bioinformática (ADN)** - Ingentes cantidades de datos – Clasificación, predicción.

# APLICACIONES COMPUTACIONALES

---

- Aplicaciones que aprenden y mejoran con la experiencia: Mozilla – spam.
- **AUTO-PROGAMACION** de robots -  
Aprendizaje mediante redes neuronales (corrección de errores de un brazo robótico industrial)
- **WEB** - Motores de búsqueda (mindset de Yahoo), navegación personalizada (hiperenlaces personalizados), tiendas virtuales (Amazon)
- **SISTEMAS ADAPTATIVOS** (e-learning).

# INDUSTRIA DEL ENTRETENIMIENTO

---

- HSS: Predicción del posible éxito de un grupo musical
  - <http://www.hitsongscience.com/>
  - Norah Jones, Maroon 5.
  - Aquí hay una forma de ganar mucho dinero.
- Pandora, Last-fm
- <http://www.music-map.com/>
- Composición musical
- Predicción de los ganadores del festival de Sundance

# OTROS

---

- Deportes - Predicción de lesiones antes de contratar a un futbolista – Redondo en el Milán.
- Policía - Detección y predicción de posibles policías corruptos

# RECONOCIMIENTO DE PATRONES

---

- Reconocimiento de voz
- Clasificación de documentos
- OCR (Reconocimiento Óptico de Caracteres)
  - Reconocimiento de escritura, firmas, talones, cheques...
- Reconocimiento de caras humanas
- Reconocimiento de huellas digitales o el iris, con aplicación en seguridad informática.

# LINGÜÍSTICA

---

- Traductores de inglés
  - ¿Cómo funciona Google translate?
- Comprensión de texto – resúmenes de prensa

# ROBOTS AUTÓNOMOS

---

- Conducción autónoma de vehículos
- Visión artificial

# Y ADEMÁS...

---

- Tengo que aprobar para ser ingeniero
- Pero eso es lo de menos

# ¿Cómo aprobar la asignatura?

---

- Estudiando mucho
- No faltando a ninguna clase
- Y haciendo muchos ejercicios