

# Introducción a la minería de datos

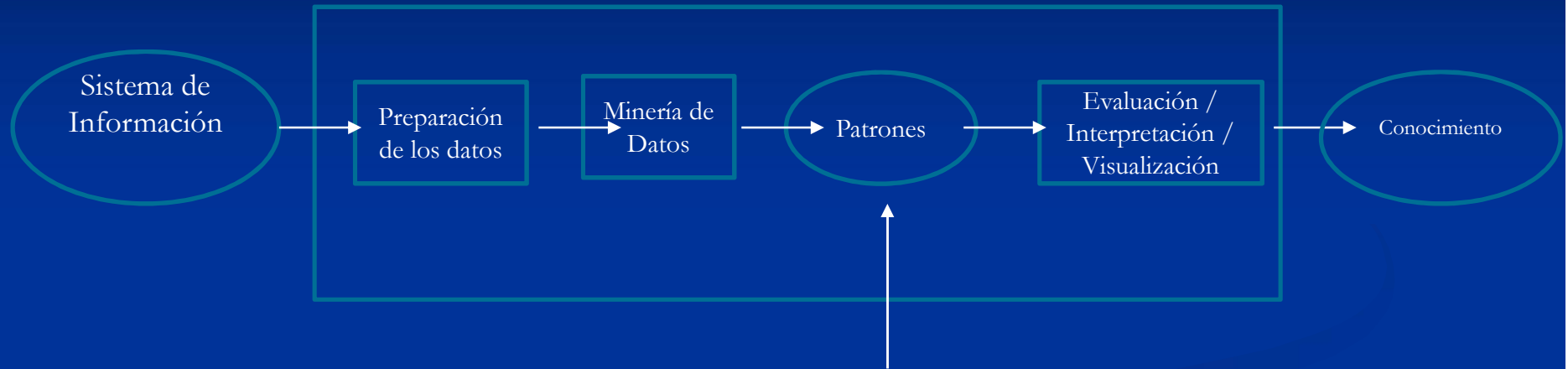
# Temario

- ¿Qué es minería de datos?
- ¿Quién usa minería de datos?
- ¿Por qué de la minería de datos?
- Ciclo virtuoso de la minería de datos

# Definición de minería de datos

- Minería de datos es la exploración y análisis de grandes cantidades de datos con el objeto de encontrar patrones y reglas significativas (conocimiento)

# Definición de minería de datos

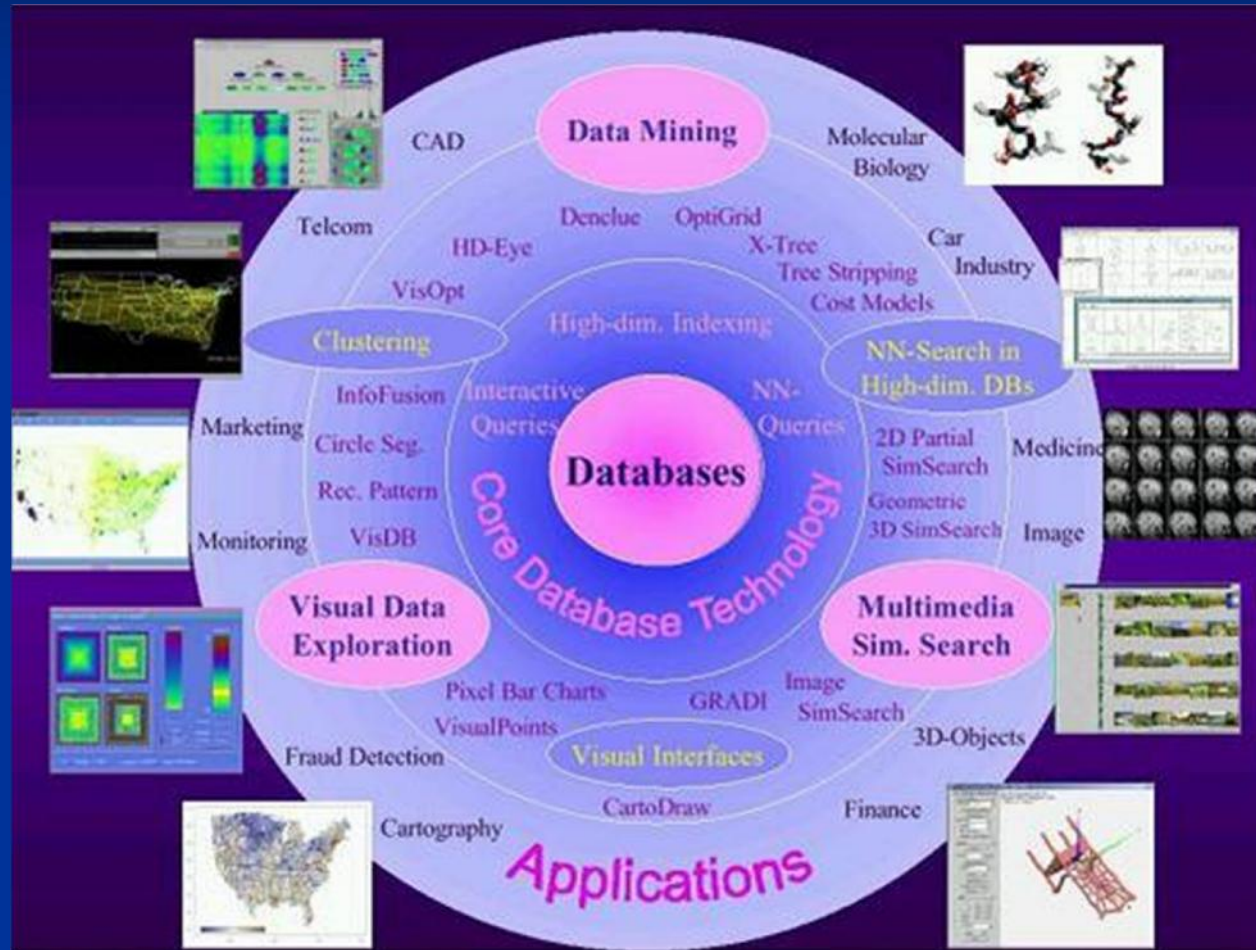


La minería de datos se refiere a la aplicación de métodos de aprendizaje y estadísticos para la obtención de patrones y modelo

# La minería de datos es un campo multidisciplinario



# La minería de datos es un subconjunto de la inteligencia de negocios



# Minería de datos

- Proceso de utilizar datos “crudos” para inferir importantes relaciones entre ellos
- Colección de técnicas poderosas para analizar grandes volúmenes de datos
- No existe un solo enfoque para minería de datos sino un conjunto de técnicas que se pueden utilizar de manera independiente o en combinación
- Existe una relación con la estadística, aunque frecuentemente se separan las técnicas que no están basadas en métodos estadísticos

# Tipos de aplicaciones de la minería de datos

- Aplicaciones o problemas de minería de datos pueden clasificarse en las siguientes categorías
  - Clasificación
  - Estimación
  - Pronóstico
  - Asociación
  - Agrupación o segmentación



# Clasificación

- Examinar las características de un nuevo objeto y asignarle una clase o categoría de acuerdo a un conjunto de tales objetos previamente definido
- Ejemplos:
  - Clasificar aplicaciones a crédito como bajo, medio y alto riesgo
  - Detectar reclamos fraudulentos de seguros

# Estimación

- Relacionado con clasificación
  - Mientras clasificación asigna un valor discreto, estimación produce un valor continuo
- Ejemplos:
  - Estimar el precio de una vivienda
  - Estimar el ingreso total de una familia

# Pronóstico

- Predecir un valor futuro con base a valores pasados
- Ejemplos:
  - Predecir cuánto efectivo requerirá un cajero automático en un fin de semana

# Asociación

- Determinar cosas u objetos que van juntos
- Ejemplo:
  - Determinar que productos se adquieren conjuntamente en un supermercado

# Agrupación o segmentación

- Dividir una población en un número de grupos más homogéneos
- No depende de clases pre-definidas a diferencia de clasificación
- Ejemplo:
  - Dividir la base de clientes de acuerdo con los hábitos de consumo

# Usos de la minería de datos

- Administración de la relación con los clientes:
  - Identificar nuevos clientes potenciales para aumentar ventas
  - Ampliar la base de cliente con la mínima inversión por parte de la empresa
  - Retener clientes existentes evitando que se vayan a la competencia (“attrition”)
  - Vender más a clientes existentes (“ventas cruzadas”)

# Usos de la minería de datos

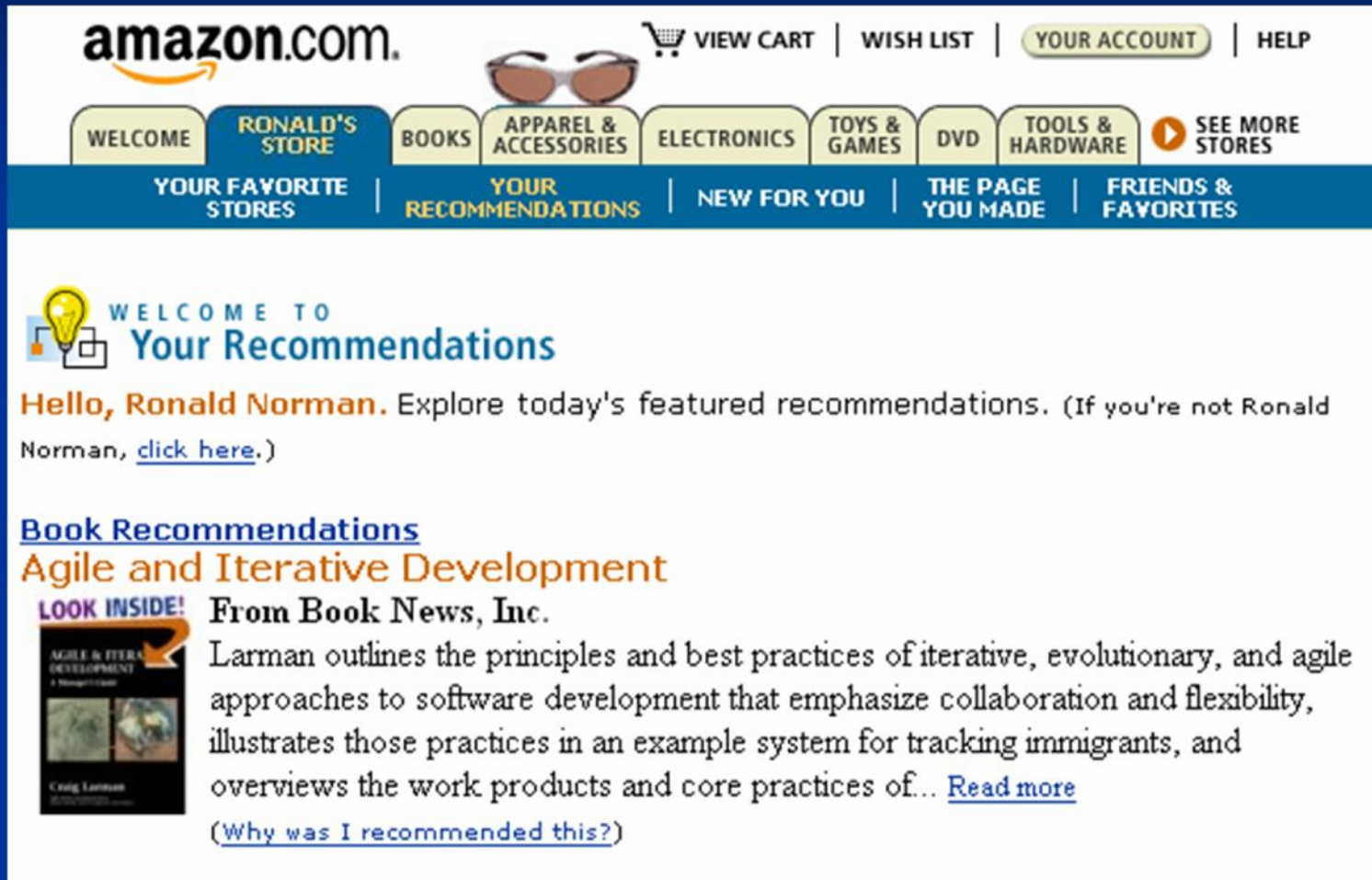
- Detección de fraudes en el uso de tarjetas de crédito
- Determinar patrones que puedan estar relacionados con lavado de dinero
- Determinar el precio de una casa con base en sus características y el precio de otras casas vendidas

# Usos de la minería de datos


- Usos de la minería de datos se han ampliado con el comercio electrónico



# Recomendaciones




The image shows a screenshot of the Amazon.com website. At the top, the Amazon logo is on the left, and navigation links for 'VIEW CART', 'WISH LIST', 'YOUR ACCOUNT', and 'HELP' are on the right. Below this is a category bar with buttons for 'WELCOME', 'RONALD'S STORE', 'BOOKS', 'APPAREL & ACCESSORIES', 'ELECTRONICS', 'TOYS & GAMES', 'DVD', 'TOOLS & HARDWARE', and 'SEE MORE STORES'. A secondary bar contains 'YOUR FAVORITE STORES', 'YOUR RECOMMENDATIONS', 'NEW FOR YOU', 'THE PAGE YOU MADE', and 'FRIENDS & FAVORITES'. The main content area features a 'WELCOME TO Your Recommendations' header with a lightbulb icon. Below it, a personalized message says 'Hello, Ronald Norman. Explore today's featured recommendations. (If you're not Ronald Norman, [click here.](#))'. A section titled 'Book Recommendations' highlights 'Agile and Iterative Development' from Book News, Inc. It includes a book cover image and a description: 'Larman outlines the principles and best practices of iterative, evolutionary, and agile approaches to software development that emphasize collaboration and flexibility, illustrates those practices in an example system for tracking immigrants, and overviews the work products and core practices of... [Read more](#)'. A link '(Why was I recommended this?)' is also present.

amazon.com.  VIEW CART | WISH LIST | YOUR ACCOUNT | HELP

WELCOME | RONALD'S STORE | BOOKS | APPAREL & ACCESSORIES | ELECTRONICS | TOYS & GAMES | DVD | TOOLS & HARDWARE | SEE MORE STORES


YOUR FAVORITE STORES | YOUR RECOMMENDATIONS | NEW FOR YOU | THE PAGE YOU MADE | FRIENDS & FAVORITES

 WELCOME TO  
**Your Recommendations**

**Hello, Ronald Norman.** Explore today's featured recommendations. (If you're not Ronald Norman, [click here.](#))

**Book Recommendations**  
**Agile and Iterative Development**

**LOOK INSIDE!** From Book News, Inc.

 Larman outlines the principles and best practices of iterative, evolutionary, and agile approaches to software development that emphasize collaboration and flexibility, illustrates those practices in an example system for tracking immigrants, and overviews the work products and core practices of... [Read more](#)

([Why was I recommended this?](#))

# Disponibilidad de datos de transacciones

The screenshot shows the Amazon.com website interface. At the top, there is a navigation bar with the Amazon logo, a shopping cart icon, and links for 'VIEW CART', 'WISH LIST', 'YOUR ACCOUNT', and 'HELP'. Below this is a secondary navigation bar with 'WELCOME', 'RONALD'S STORE', and various product categories: 'BOOKS', 'APPAREL & ACCESSORIES', 'ELECTRONICS', 'TOYS & GAMES', 'MUSIC', and 'BABY'. A 'SEE MORE STORES' button is also present. The main content area shows the breadcrumb path: 'Account > Where's My Stuff? > Orders placed in 2004'. Below this is a 'See more' section with a dropdown menu set to '-Select different orders to view-' and a 'GO!' button. A help link 'Need help using this page?' is also visible. The 'Your Orders' section lists three orders from 2004, each with a 'View order' button. The first order is from Mar 16, 2004, for 'Balancing Agility and Discipline: A Guide for the Perplexed'. The second is from Feb 15, 2004, for 'Test Driven Development: By Example [Paperback] by Beck, Kent'. The third is from Feb 11, 2004, for 'Extreme Programming Explained: Embrace Change [Paperback] by Beck, Kent'.

Shop in **Sports & Outdoors** Beta-What is this?

amazon.com. VIEW CART | WISH LIST | YOUR ACCOUNT | HELP

WELCOME RONALD'S STORE BOOKS APPAREL & ACCESSORIES ELECTRONICS TOYS & GAMES MUSIC BABY SEE MORE STORES **Ronald's G**

Account > Where's My Stuff? > Orders placed in 2004

See more  [Need help using this page?](#)

### Your Orders

**Order Date:** Mar 16, 2004 **View order**  
**Order #:** 002-0135642-1254476  
**Recipient:** Ronald Norman  
**Items:**

- 1 of Balancing Agility and Discipline: A Guide for the Perplexed

---

**Order Date:** Feb 15, 2004 **View order**  
**Order #:** 058-5303369-6295505  
**Recipient:** Ronald Norman  
**Items:**

- 2 of Test Driven Development: By Example [Paperback] by Beck, Kent

---

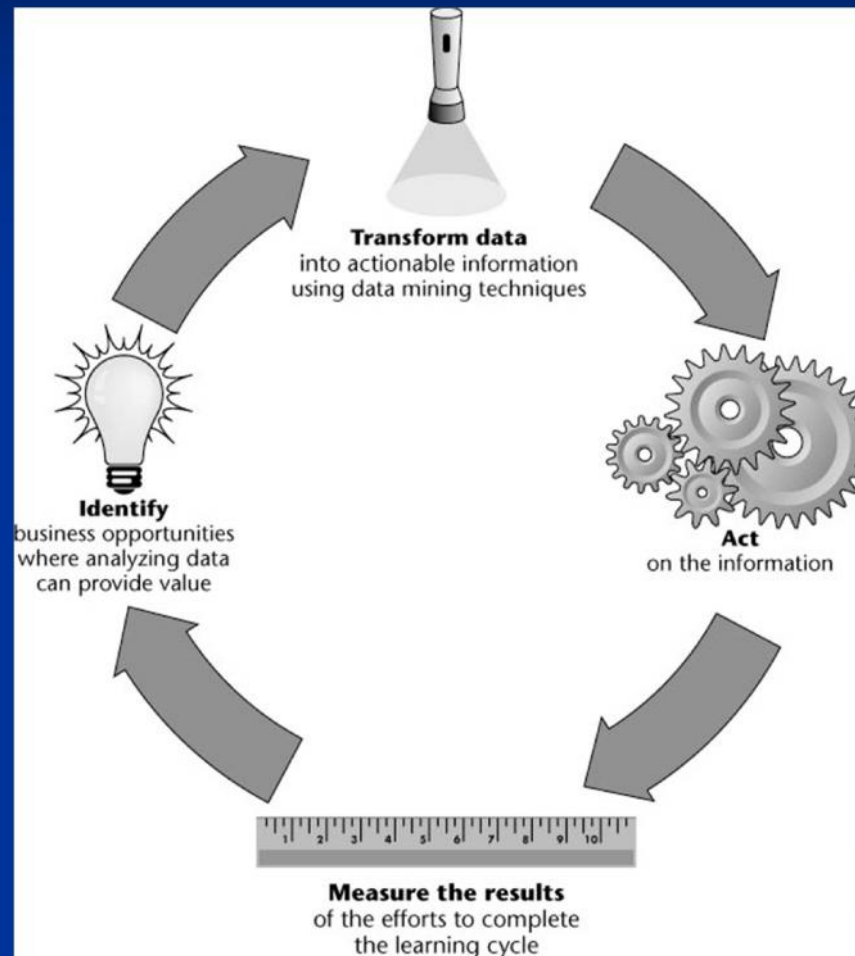
**Order Date:** Feb 11, 2004 **View order**  
**Order #:** 058-7996307-9045133  
**Recipient:** Ronald Norman  
**Items:**

- 2 of Extreme Programming Explained: Embrace Change [Paperback] by Beck, Kent

# ¿Por qué de la minería de datos?

- Datos se encuentran disponibles
- Poder computacional es cada vez menos costoso
- Las presiones competitivas son enormes
- Software para minería de datos se encuentra disponible

# Ciclo virtuoso de la minería de datos



# Principales técnicas de minería de datos

- Análisis de encadenamiento
- Árboles de decisión
- Redes neuronales artificiales
- Algoritmos genéticos