

**INVERSIONES  
CARROUSEL**  
S.A. DE C.V.

**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones



**Minería de Datos Para la Inteligencia de Negocios**

Luis Francisco Zaldivar, MSE  
[www.modelacionderiesgos.com](http://www.modelacionderiesgos.com)

**INVERSIONES  
CARROUSEL**  
S.A. DE C.V.

**ORACLE**  
CRYSTAL BALL

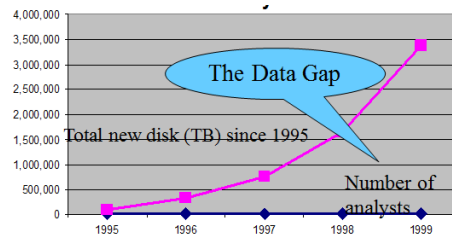
Modelación de riesgos para la toma de decisiones

**Porqué el Análisis de Datos Tradicional es Limitado?**

1. Tremenda y creciente cantidad de información
2. Múltiples fuentes de información
3. Información es multidimensional
4. Cada día la información es más complicada
5. Nuevas y sofisticadas aplicaciones
6. Información es insumo de Modelos de Simulación y Optimización Estocástica

### Brecha Data Versus Analistas

1. Analistas requieren mucho tiempo para tratar de analizar la información
2. Existe muchos patrones en la información escondidos que no son fáciles de describir y/o detectar
3. Mucha información nunca se analiza!



### Beneficios

1. Incrementar Rentabilidad
2. Reducir Riesgos
3. Facilitar Cumplimiento de Metas

**INVERSIONES**  
**CARROUSEL**  
S.A.DE C.V.

**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones

## Definiciones

1. Capacidad de extraer información importante para la toma de decisiones
2. Proceso de explorar y analizar, por medios automáticos o semiautomáticos, bases de datos grandes para poder descubrir patrones importantes y reglas
3. Proceso importante para determinar correlaciones, patrones y tendencias en las bases de datos a través del uso de técnicas estadísticas y matemáticas

**INVERSIONES**  
**CARROUSEL**  
S.A.DE C.V.

**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones

## Alcance de la Minería de Datos

La Minería de Datos es aplicada a una muestra importante de la Base de Datos de la Empresa, para luego después de seleccionar la técnica mas adecuada, proceder a calificar la base de datos no empleada. Esto facilitará la toma de decisiones que creen valor económico a través de subir la rentabilidad y reducir el riesgo en la operación de la empresa.

**INVERSIONES**  
**CARROUSEL**  
S.A.DE C.V.

**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones

## Variables a Analizar

1. Principalmente: Categóricas vs. Numéricas
2. Numérica
  1. Continuas
  2. Enteras
3. Categóricas
  1. Ordenadas (bajo, medio, alto)
  2. No Ordenadas (hombre, mujer)

**INVERSIONES**  
**CARROUSEL**  
S.A.DE C.V.

**ORACLE**  
CRYSTAL BALL

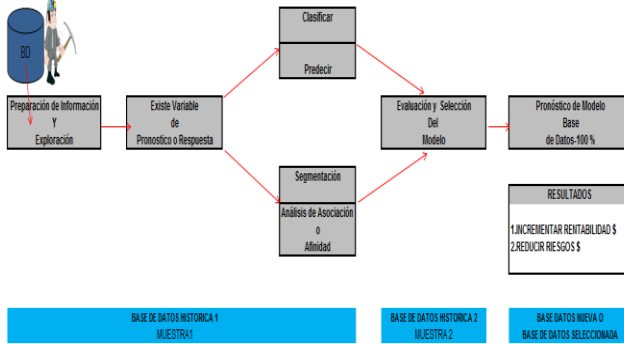
Modelación de riesgos para la toma de decisiones

## Tratamiento de la Muestra en la Base de Datos

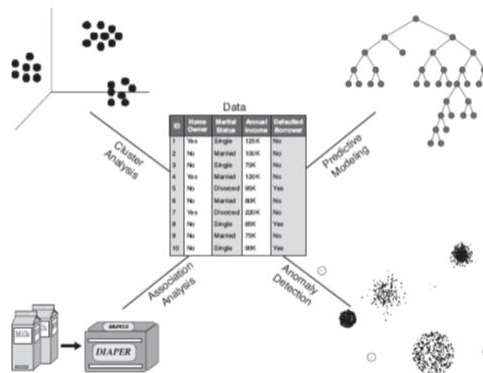
1. Muestreo de cualquier base de datos, incluyendo MS PowerPivot en memoria, base de datos que manejan hasta 100 millones de filas.
2. Limpieza de datos a través de categorizar la base de datos y rellenar datos faltantes
3. Partición de la base de datos a través de Base Modelo, Base Validación y Base de Prueba

### Proceso

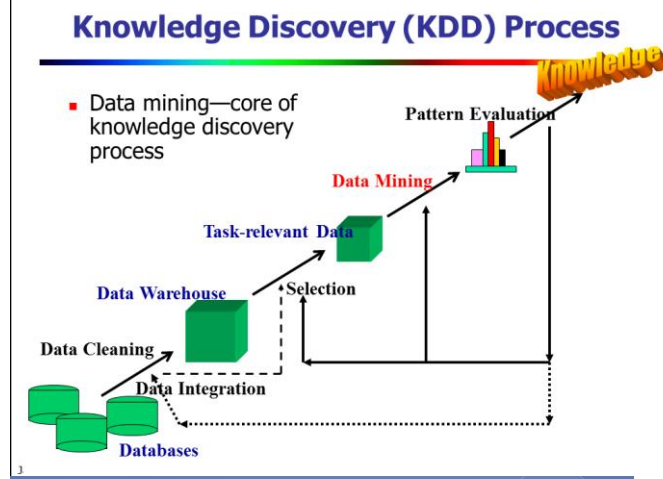
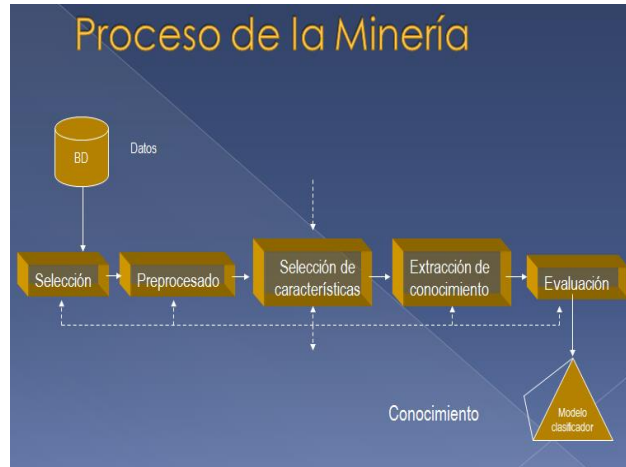
PROCESO DE MINERIA DE DATOS PARA LA INTELIGENCIA DE NEGOCIOS



### Trabajos Básicos



Four of the core data mining tasks.



**INVERSIONES CARROUSEL**  
S.A. DE C.V.

**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones

## Proceso

1. Definir/Entender el propósito del trabajo
2. Obtener Datos de la Base (puede involucrar muestreo)
3. Explorar, limpiar y pre-procesar los datos
4. Reducir datos a analizar
5. Especificar trabajo a realizar: Clasificar & Predecir, Asociatividad)
6. Seleccionar técnica a emplear
7. Usar y aplicar todas las técnicas disponibles
8. Analizar los resultados y compararlos con los modelos disponibles
9. Seleccionar el mejor Modelo
10. Aplicarlo a la base de datos
11. Tomar decisiones

**INVERSIONES CARROUSEL**  
S.A. DE C.V.

**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones

## Análisis con Variable de Pronóstico

1. Meta: Pronosticar Variable Categórica del Cliente
2. Ejemplo: Compra/no Compra, fraude/no fraude, sujeto de crédito/no sujeto de crédito...
3. Cada fila es un caso (cliente)
4. Cada columna es una variable( Ingreso, sexo)
5. Usualmente la Variable de Pronóstico es Binaria (si/no)

## Análisis sin Variable de Pronóstico

1. Meta: Segmentar datos y crear reglas que definan "Qué va con qué". Ejemplo: "Si X compró, Y también compró"
2. Las filas son transacciones
3. Usados para recomendar o aconsejar a clientes - "Nuestra información muestra que usted compró el producto X, podría comprar también el producto Y"
4. Se conoce como "Análisis de Afinidad"
5. Importante para Identificar Patrones

## Técnicas Disponibles Cuando Existe Variable de Pronóstico



Técnicas Predicción
1. Regresión Múltiple 2. K-Mi Vecino Cercano 3. Árboles de Regresión 4. Redes Neuronales 5. Simulación
Clasificación
1. K-Vecino mas Cercano 2. Naive Bayes 3. Árboles de Clasificación 4. Análisis Discriminante 5. Regresión Logística 6. Redes Neuronales
Pronósticos con Series de Tiempos
1. Métodos de Regresión 2. Series Temporales




**INVERSIONES CARROUSEL**  
S.A. DE C.V.

**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones

## Técnicas Disponibles para Análisis de Afinidad o Asociatividad



Análisis de Afinidad
1. Reglas de Asociatividad
2. Analisis Cluster
3. Componentes Priniopales

**INVERSIONES CARROUSEL**  
S.A. DE C.V.

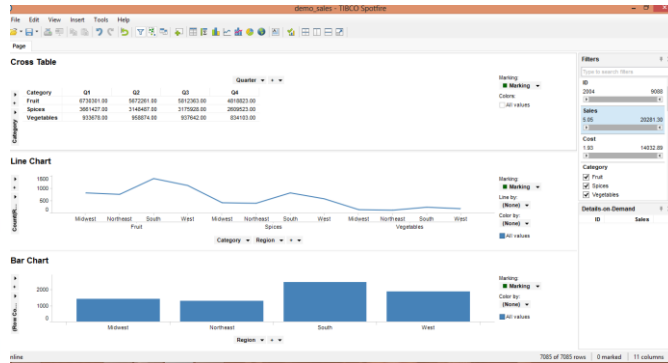
**ORACLE**  
CRYSTAL BALL

Modelación de riesgos para la toma de decisiones

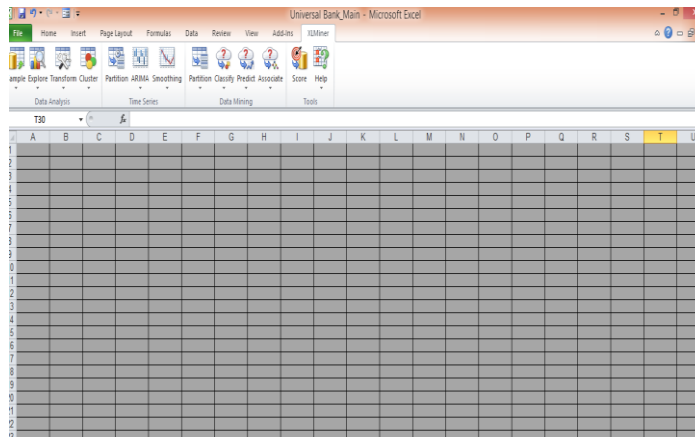
## Tecnologías a emplear

1. MS Excel®
2. XLMiner®
3. Oracle Crystal Ball®
4. TIBCO Spotfire®

## Barra de Herramientas TABCO Spotfire

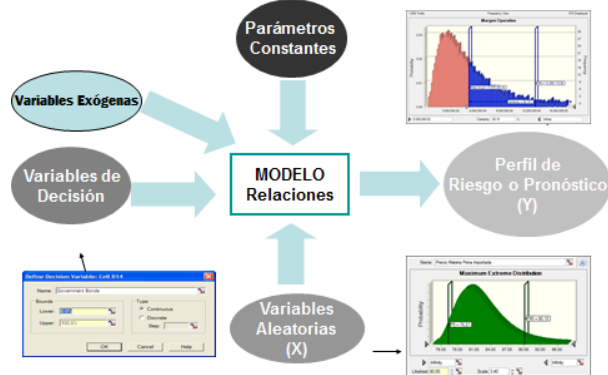


## Barra de Herramientas de XLMiner



## Mapa de Trabajo Oracle Crystal Ball

### Modelo de Análisis de Riesgo



### Otros beneficios

1. Ambiente de MS Excel que facilita el aprendizaje de la temática
2. Modelos desarrollados sirven para poder calificar bases de datos grandes
3. XLMiner tiene funciones para interactuar con varias bases de datos (tomar muestras de las bases de datos y establecer pronósticos con el modelo desarrollado)
4. XLMiner interactúa con -- PowerPivot, Bases de Datos: Microsoft/IBM/Oracle y hoja de cálculo
5. Capacidad de MS Excel 2010: 1 millón de filas x 16,000 columnas.