

A conceptual illustration of a man in a white shirt and tie holding a glowing globe. The background is filled with various symbols representing data and finance, including binary code (0s and 1s), a bar chart, a globe, and various letters and numbers. The overall color palette is a mix of blue, green, and yellow.

# Los Modelos de Ventas Cruzadas en la Industria Financiera (Cross-Sell/Upsell Solution for Banking)

---

Viterbo H. Berberena González  
Doctor en Ciencias Técnicas  
Mayo 2004

# Personal Involucrado

- El personal de negocio.
  - Cruzar la venta de tarjetas créditos y productos relacionados con los poseedores de otros productos.
  - Vender productos o servicios relevantes a los tenedores de tarjetas de crédito.
  - Utilizar la información analítica para segmentar los clientes y administrar las campañas.
  
- El personal analítico.
  - Desarrollar modelos dentro de un proyecto de minería de datos predefinido y flexible.
  - Crear un paquete de resultados de los modelos, para que sea revisado por los usuarios de negocio.
  - Integrar los resultados y los modelos en la asignación de puntajes y el desarrollo de las campañas.

# Componentes

- Data mart analítico.
- Interfase de extracción de datos.
- Integración de las fuentes de datos externas.
- Desarrollo del modelo para medir la propensión a las ventas cruzadas con el minero de datos.
- Interfase para sacar los resultados y el diagnóstico.
- Registro del modelo en el data mart para la asignación subsecuente de los puntajes (scoring).
- Personalizar los requerimientos especiales para los mercados globales.

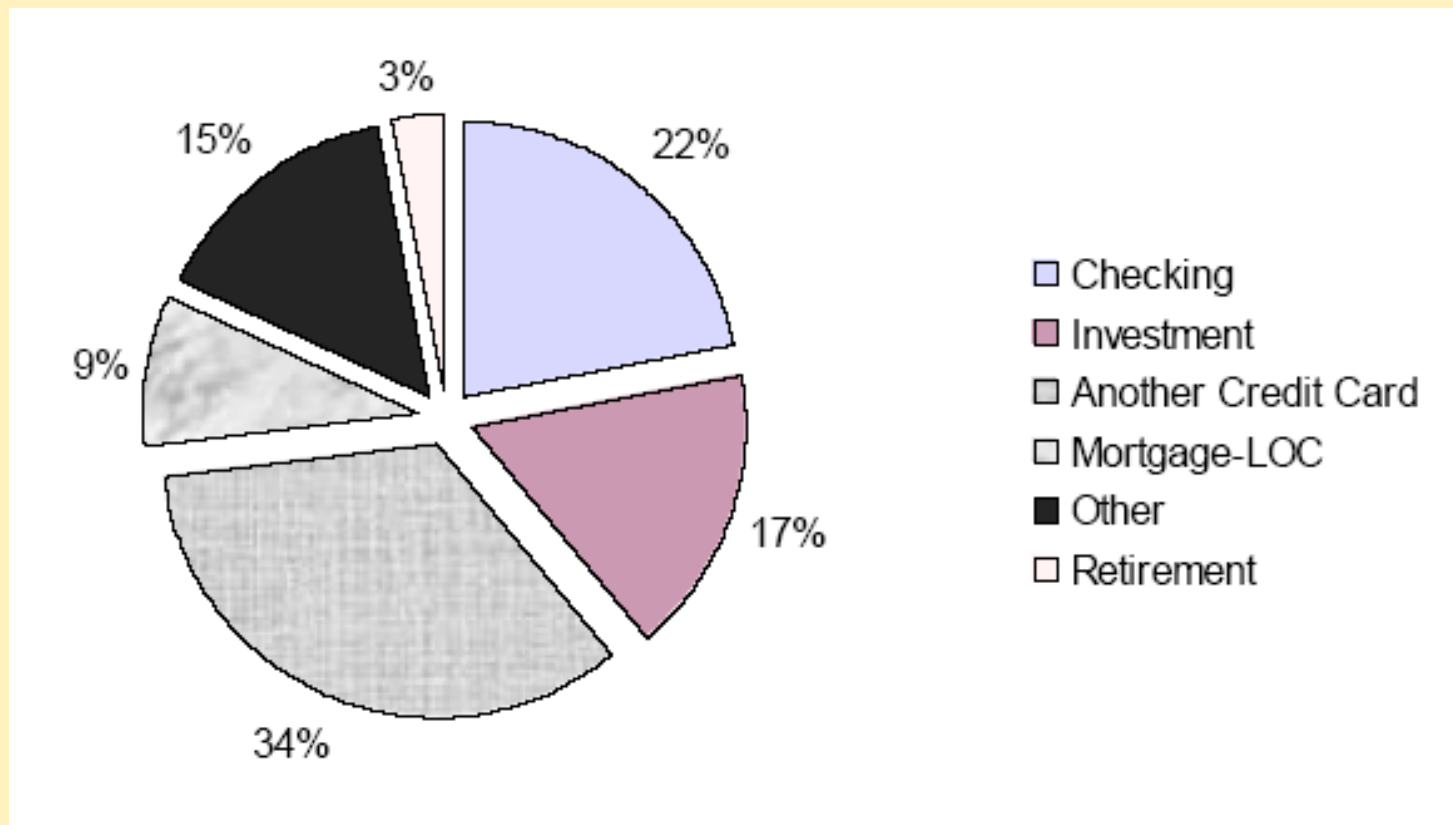
# Alcance

- Delimitación del problema de negocio y de la variable objetivo.
- Definición de las variables independientes (predictors).
- Acondicionamiento de los datos en el data mart.
- Especificaciones del modelo, como la estructura y las características técnicas analíticas.
- Utilización, incluyendo la asignación de puntajes (scoring) y el informe de resultados.

# El Problema de Negocio

- Utilizar las relaciones con los clientes para aumentar las ventas.
- Cruzar las ventas de diferentes líneas de productos.
  - Tarjetas de crédito a clientes de inversiones.
- Vender productos o servicios adicionales (de valor agregado) dentro las líneas existentes.
  - Protección al crédito en las tarjetas de crédito.
- Subir la categoría de los productos dentro de las líneas existentes.
  - De tarjeta clásica a oro, de oro a platino, etc.

# Cross-Sell Sources for Credit Cards



# Formulación del Problema de Negocio

- La meta es determinar la propensión a la respuesta y aceptación de las ofertas de tarjetas de crédito, a través de un canal directo, dentro de un temporalidad predefinida (Cross-Sell).
- Los datos actuales de respuesta se necesitan porque las ofertas relacionadas con crédito son sensibles a la temporalidad (al momento específico), al riesgo y a los requerimientos regulatorios.
- Adicionalmente se obtiene la propensión de ventas de productos y servicios adicionales, de valor agregado, dentro de las líneas de productos de tarjetas de crédito, así como el aumento de categoría de los productos existentes (Up-Sell / Add-on / Upgrade).

# Definición de la Variable Objetivo

- Criterio Primario (target definition):
  - Poseedores de tarjetas de crédito (clientes a los que se le a ofertado y han aceptado un producto de tarjeta de crédito dentro de un período de tiempo reciente: 6 – 12 meses).
  - Otros productos de tarjetas de crédito (clientes que actualmente poseen productos de mayor categoría).
- Criterio Secundario (target subdefinition):
  - Titularidad de los poseedores de tarjetas (titulares con al menos 6 meses de antigüedad).
  - Historial de uso continuo (uso continuo en ciclos previos de facturación: 3 – 6 ciclos).
  - Tipos de saldos (seleccionar de pagos regulares, revolventes o saldos inactivos).
  - Rentabilidad o ingresos de los últimos 6 meses por clientes (rentabilidad, ingresos o intereses/cargos financieros > 0).

## Aspectos importantes sobre los filtros para la refinación del modelo

- Los criterios adicionales ayudan a determinar los clientes más deseables a ser abordados.
- Limitar los datos para la modelación significa que, el modelo está constreñido a una subpoblación más pequeña, pero es mucho más preciso.
- Un número de alternativas por omisión, que se incorporan en el modelo serán validadas o modificadas por el usuario, antes de correr el modelo.
- La refinación de los criterios utilizados para la definición de la problema pueden ser manejados por el marketing (consumidores) o por valoraciones de riesgo (actuarial).
- Este enfoque integrado optimiza ambos criterios antes de que se inicien las campañas de mercadotecnia.

# Filtros para la refinación de los criterios

- Aplicación de límites de riesgo (límites de los puntajes de crédito internos).
- Puntaje del Buró de Crédito (Puntaje externo dado por el Buró de Crédito).
- Límite 1: Proporción del servicio de la deuda total (porcentaje del límite de crédito con respecto a la deuda total cliente).
- Límite 2: Proporción del servicio de la deuda neta (desembolso de efectivo en los pagos mensuales).
- Límite 3: Número total de tarjetas externas (número de tarjetas activas en el Buró de Crédito).
- Riesgo de concentración (magnitud del riesgo que enfrenta la institución).
- Taza de aprobación (respuesta a la tasa de aceptación por parte del segmento de clientes).

# Horizontes de Tiempos en la Selección de los datos

- La selección del horizonte de tiempo (para la modelación) y los datos, ayudan a responder dos preguntas cruciales:
  - ¿Qué nivel del horizonte de tiempo se desea modelar?
  - ¿Qué grupo de clientes está buscando para cruzar ventas con tarjetas de crédito?
- El horizonte de tiempo tiene dos dimensiones:
  - La frecuencia transaccional, usada para determinar cuánto de la profundidad inherente al comportamiento transaccional debe ser capturado.
  - El horizonte de tiempo es también determinado como el período de tiempo sobre el cual los datos, acerca de las variables independientes, debe ser recogidos.
  - Ambas dimensiones tienen un impacto substancial en la calidad y la dirección del modelo.

## Aspectos importantes sobre la población objetivo para la modelación

- Los modelos de ventas cruzadas son más efectivos cuando se desarrollan sobre los clientes, a quienes se quiere dirigir las tarjetas de crédito.
- Se puede obtener la propensión a las ofertas de tarjetas de crédito para todos los clientes, con un sólo modelo. En este caso la población general objetivo puede ser todos los clientes que en la actualidad poseen tarjetas de crédito.
- El modelo es más preciso, cuando se desarrolla sobre una población más pequeña y homogénea, como por ejemplo, los clientes poseedores de productos de jubilación (afores).

## Criterios para la selección de los datos

- Cualesquiera de las cuentas abiertas por el titular (especificar aquella cuenta que forzosamente debe estar abierta cierto período de tiempo, o usar al titular como cliente).
- Cierre de cualesquiera otras cuentas (cantidad de cuentas, tipo y temporalidad de las clausuras).
- Niveles del horizonte de tiempo (mensualmente para reflejar los ciclos de facturación).
- Requerimientos de los datos históricos (fijar la ventana de tiempo para los datos transaccionales: 1 ~ 2 años pasados).
- Profundidad histórica de los datos (crear variable con un corrimiento hacia atrás de hasta 3, 6, 12 meses de historia).
- Universo objetivo para cruzar la venta de tarjetas (cualquier grupo preseleccionado, como por NSE, etc.).

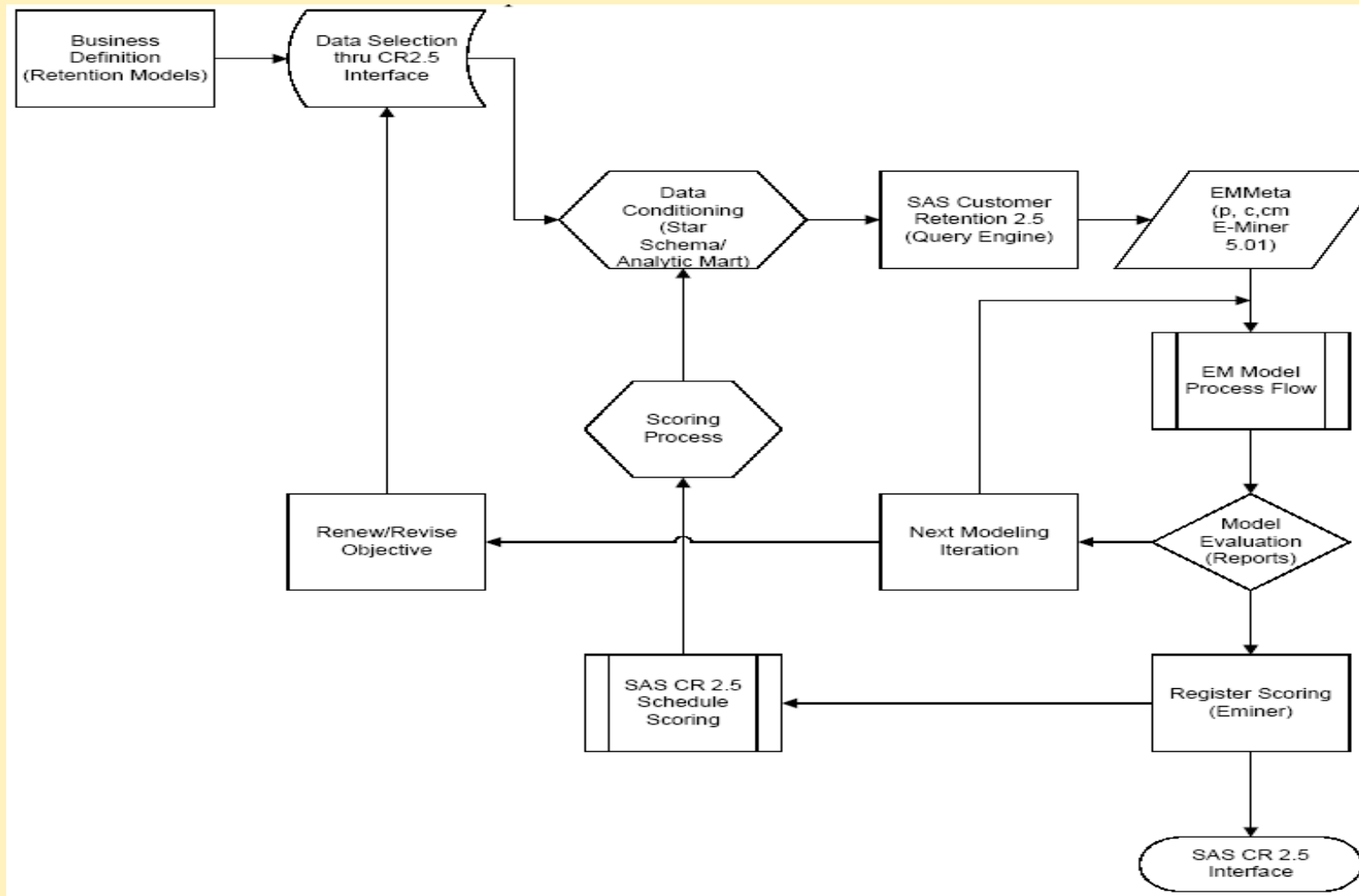
## Casos que deben ser excluidos de manera general

- Eliminación de algunas observaciones, no aprovechables por razones no analíticas, lo que resulta en aumento de la precisión de los datos, necesarios para la modelación.
- Algunos de los tipos de información (registros) que son generalmente excluidos:
  - Do not mail.
  - Fallecido.
  - Con actividad fraudulenta y otra actividad indeseable.
  - Otras circunstancias especiales que sean requeridas por el modelo.

# Metodología de Trabajo

- Creación del data mart con variables extraídas del DDS (detail data store).
- Definición del universo de clientes objetivo, por parte de los expertos en técnicas analíticas y los de negocio.
- Extracción de las observaciones que serán exportadas al minero de datos y que constituirán el conjunto de datos analíticos (analytic data set), a través de una interfase.
- Transformación de algunas variables y creación de otras.
- Desarrollo de los modelos y análisis de los resultados en el minero.
- Exportación automática de las tablas con la asignación de puntajes (scoring tables) y los parámetros del propio modelo (model score) al data mart.
- Utilización de la información para la segmentación, creación de reportes y ejecución de las campañas a través de una interfase.

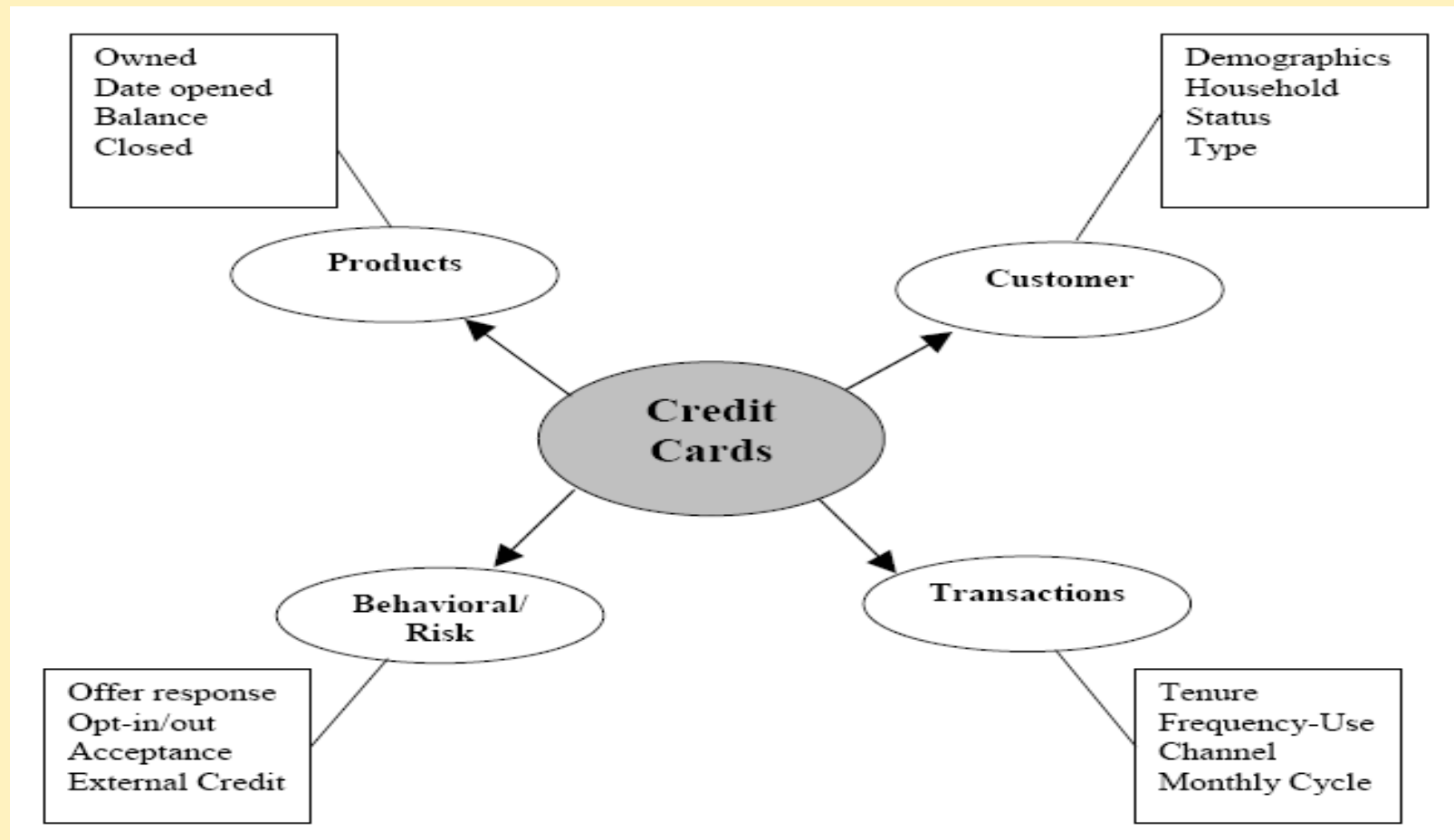
# Arquitectura de la Solución



# Composición de los Datos

- Información del cliente.
  - Demográficos, doméstico (familiar), nivel socio-económico, tipo o clase de cliente.
- Transaccionalidad.
  - Titularidad, frecuencia (uso), canal, ciclo mensual.
- Comportamiento/Riesgo.
  - Respuesta a las ofertas, abandono/regreso, aceptación, crédito externo.
- Productos.
  - Tenedor, fecha de apertura, saldo, cierre.
- Respuestas a las campañas.
  - Promociones, precios, puntos, catálogos, etc.

# Flujo de Datos Analíticos



# Variables

- Usos de las variables.
  - Filtrado, selección de datos, modelación y diagnóstico (evaluación del modelo).
- Variables independientes (predictores).
  - Extracción del data mart.
  - Preparación de datos (tipo, horizonte de tiempo, uso del modelo).
- Selección de variables.
  - Recopilación de variables de diferentes dominios (buró de crédito, medios de pagos, database marketing, etc.).
  - Evaluación (valor de uso, redundancia, comprensión, etc.).
- Creación y transformación de variables.
  - Aumento de la precisión, fiabilidad, y valor de uso.
  - Nivelación de las relaciones entre las variables.

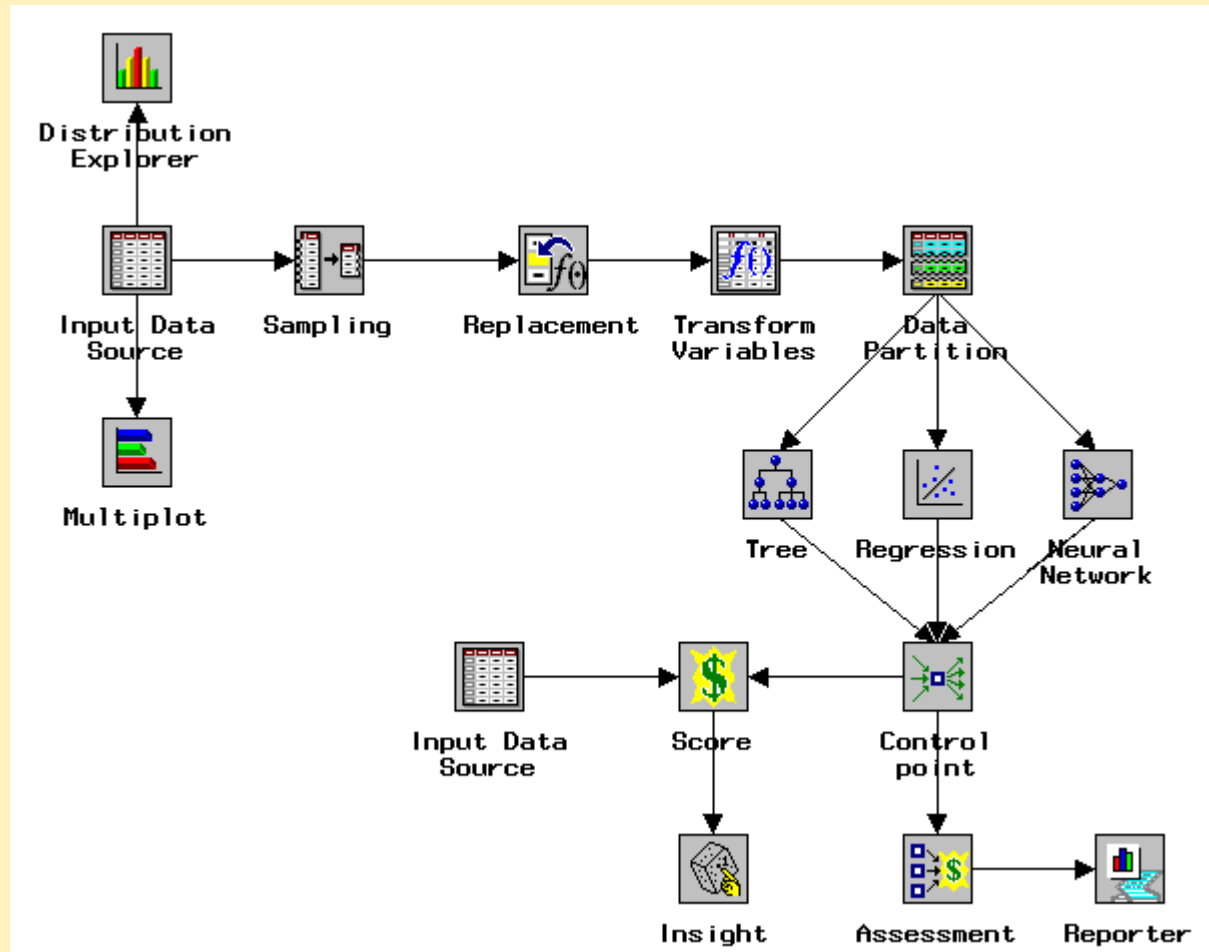
## Creación y Transformación de Variables

| Criterio  | Valor  |
|---|--|
| <ul style="list-style-type: none"> <li>■ Tipo o clase de cliente: Altos saldos, Alto NPV</li> </ul>                                       | <ul style="list-style-type: none"> <li>■ Categorizar (discretizar) las observaciones en base a la distribución.</li> </ul>   |
| <ul style="list-style-type: none"> <li>■ Horizonte de tiempo (hacia atrás) para el análisis de los saldos y las transacciones.</li> </ul> | <ul style="list-style-type: none"> <li>■ Intervalos definidos de 3, 6 y 12 meses.</li> </ul>   |
| <ul style="list-style-type: none"> <li>■ Relaciones, proporciones o razones.</li> </ul>   | <ul style="list-style-type: none"> <li>■ Saldos contra el límite, efectivo o transferencia de saldos contra saldos totales.</li> </ul>   |
| <ul style="list-style-type: none"> <li>■ Extremos (picos y depresiones).</li> </ul>   | <ul style="list-style-type: none"> <li>■ Saldo siempre alto o bajo, la mayoría de las comisiones pagadas, etc.</li> </ul>  |
| <ul style="list-style-type: none"> <li>■ Subagrupamiento.</li> </ul>  | <ul style="list-style-type: none"> <li>■ Total de efectivo y transacciones de compras por tipo de tarjeta.</li> </ul>  |
| <ul style="list-style-type: none"> <li>■ Indicadores (banderas).</li> </ul>   | <ul style="list-style-type: none"> <li>■ Siempre que se sobrepase el límite, que se liquide el saldo, que se use la protección al crédito, que se haga una solicitud para cambio de tasa o límite, etc.</li> </ul> |

# Proceso de Modelación

- **Exploración Estadística de la Base de Datos.**
  - Paquete de análisis estadístico.
  - Comprensión de las variables.
  - Revisión de escalas de medición.
  - Eliminación, en lo posible, de la redundancia y la colinealidad.
  - Análisis de la varianza de las variables (tamizado).
  - Análisis de los valores perdidos.
  - Estadígrafos descriptivos y de frecuencia, distribuciones de prob.
- **Modelación Matemática.**
  - Minero de datos.
  - SEMMA.
  - Construcción de un diagrama de proceso flexible.
  - Identificación de clientes con alta propensión a las ventas cruzadas (Cross-Sell, Add-on, Up-Sell).
  - Creación de submodelos (si es el caso): propensión a la venta cruzada, compra de nuevos productos o servicios, aumento de categoría de producto.

# Diagrama del Flujo del Minero



# Técnicas Analíticas de Modelación

- Árboles de Decisión.
  - CHAID (chi-squared automatic interaction detection).
  - CART (classification and regression trees).
  - ML (machine learning).
- Análisis de Regresión Logística.
  - Forward.
  - Backward.
  - Stepwise.
- Análisis de Regresión.
  - MPL (multilayer perceptron).
  - Ordinary RBF (radial basis function).
  - GLM (generalized linear models).

# Desempeño del Modelo

- Variables más importantes.
  - Cuentas que llevan abiertas un período de tiempo mayor que el promedio.
  - Las transacciones ocurridas en el pasado reciente que son superiores al promedio histórico.
  - Las transacciones realizadas por mes, más altas que el promedio.
  - Las tasas de interés de los clientes que son significativamente diferentes de las tasas del mercado.

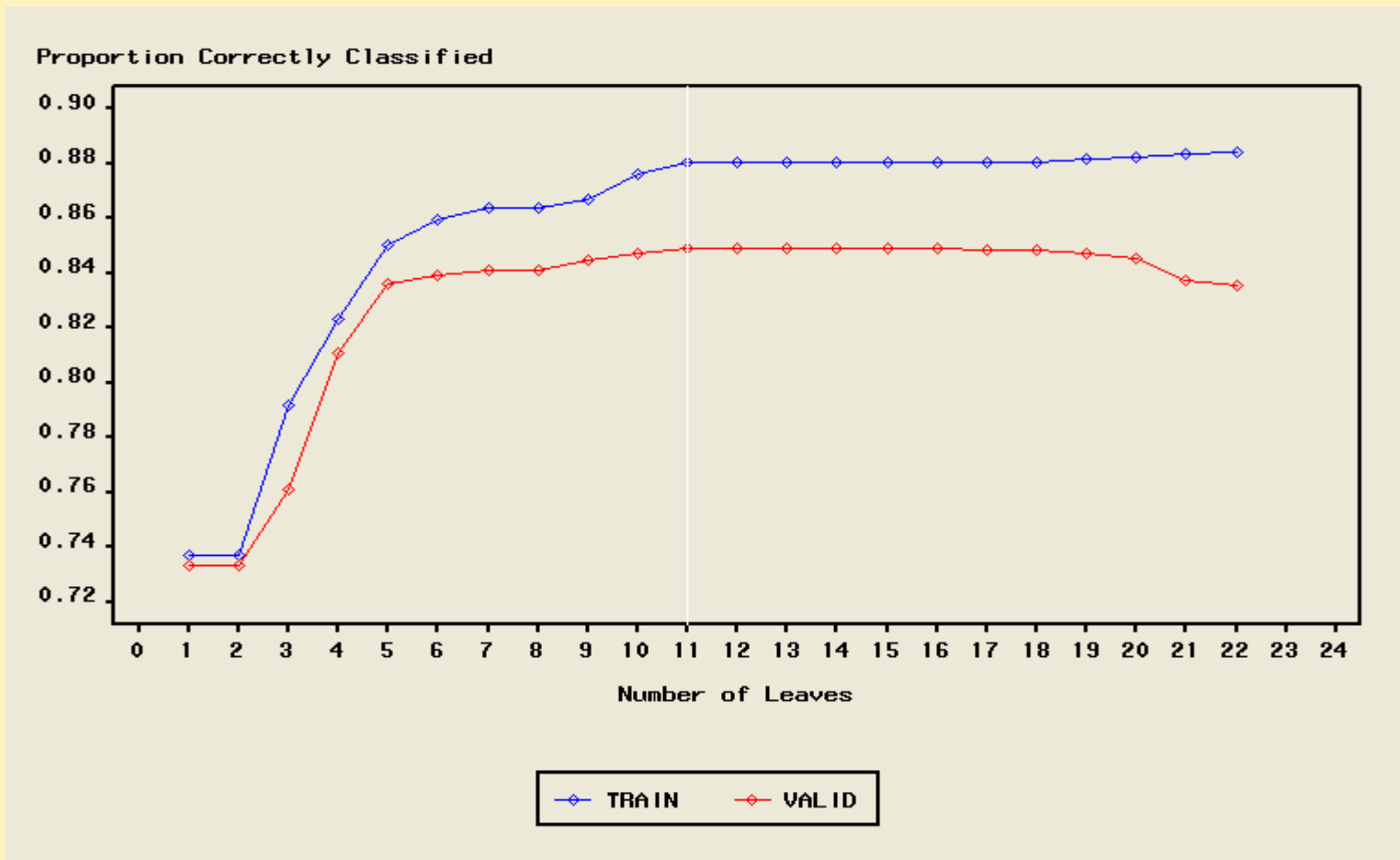
# Desempeño del Modelo

- Variables de clasificación (datos externos).
  - Las variables demográficas y las de crédito proveen bases adicionales para la segmentación en el desarrollo y aplicación del modelo, el diagnóstico y la segmentación de las campañas.
  - Variables demográficas (ingreso familiar promedio, valor de la vivienda, número de hijos, tamaño de la residencia, evaluaciones de riesgo y segmentaciones).
  - Variables de crédito (cargos totales revolventes de las cuentas, número de cuentas abiertas los últimos 12 meses, total de cuentas, evaluación de riesgo reciente, valor neto).

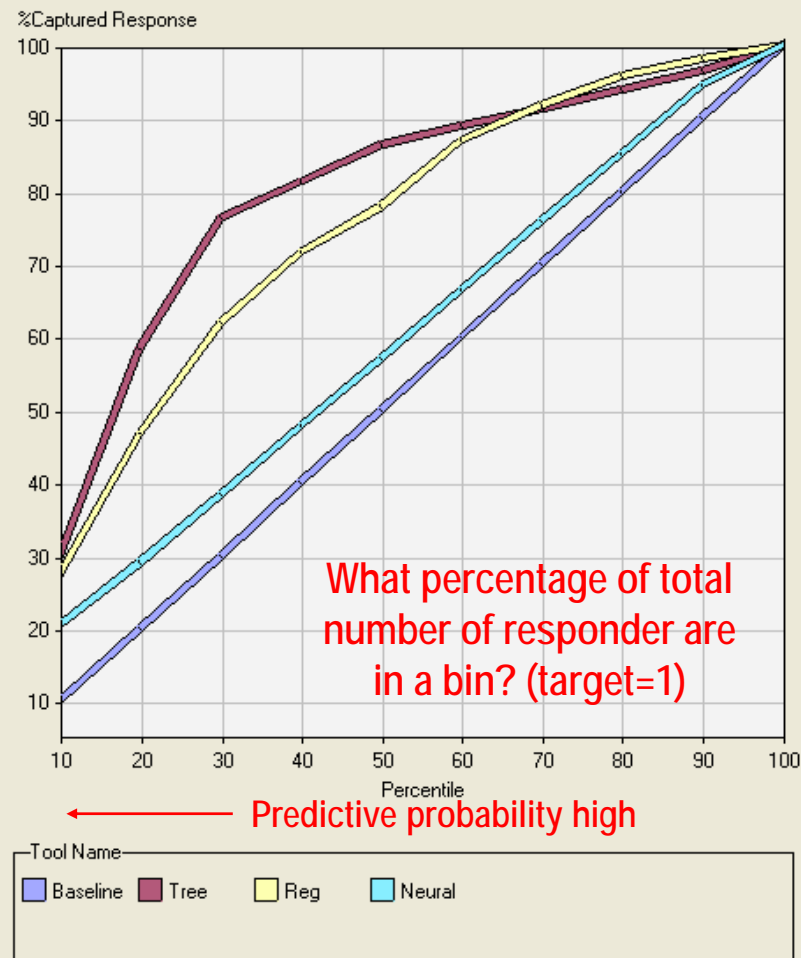
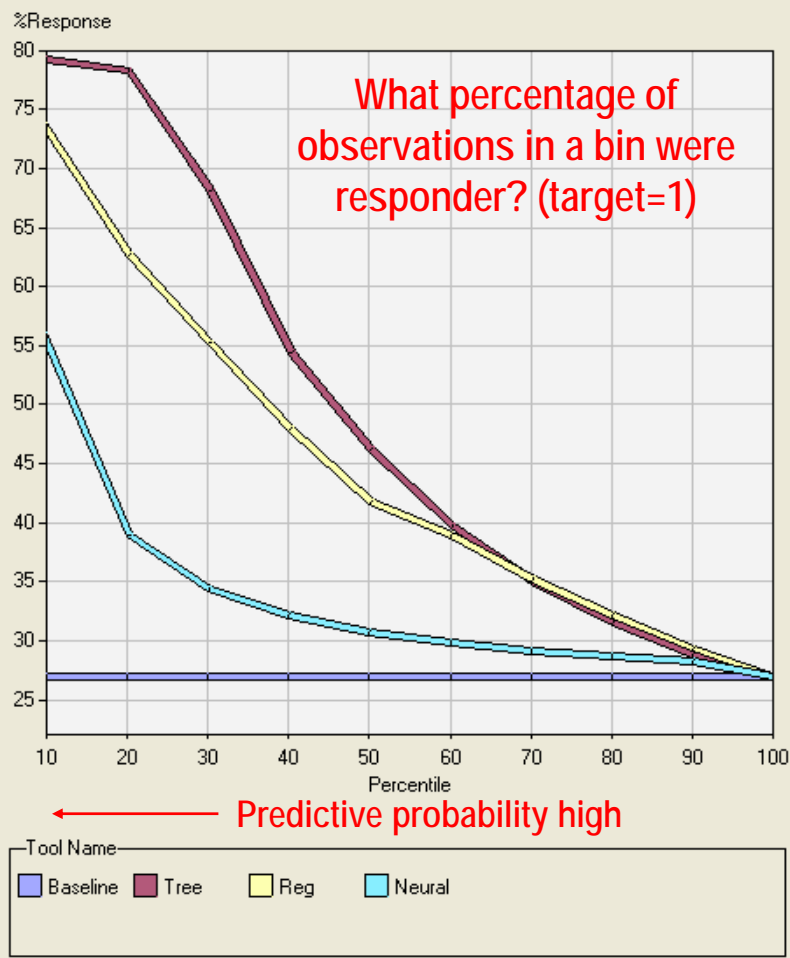
# Desempeño del Modelo

- Evaluación del desempeño.
  - Comparación de la capacidad predictiva de los diferentes tipos de modelos (lift chart).
  - Evaluación de la calidad de las estimaciones del modelo (estimaciones de la variable objetivo).
    - Habilidad para identificar la mayor propensión (incremental lift).
    - Variables que entraron al modelo.
    - Recorrido del índice lift, desde el mejor hasta el peor modelo.
    - Fortaleza de los coeficientes del modelo.
    - Consistencia del modelo con las diferentes particiones de los datos (training, validation and test).

# Training & Validation Errors (Tree)

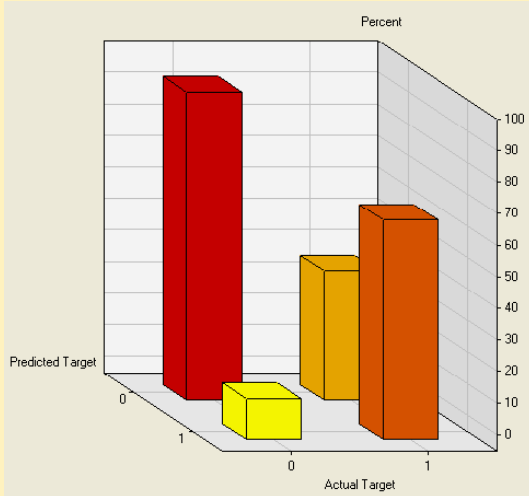


# Lift/Response Charts (cumulative)

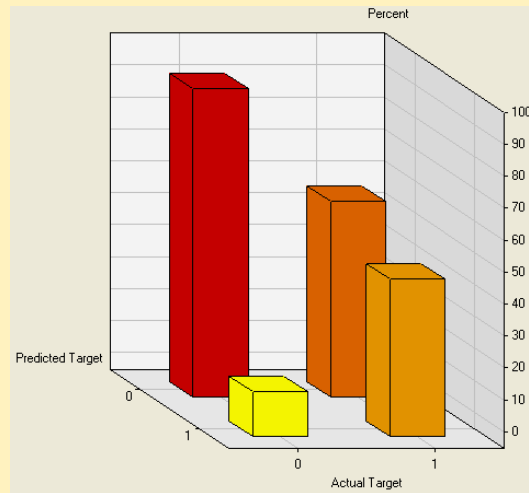


# Confusion Matrix (predictive accuracy)

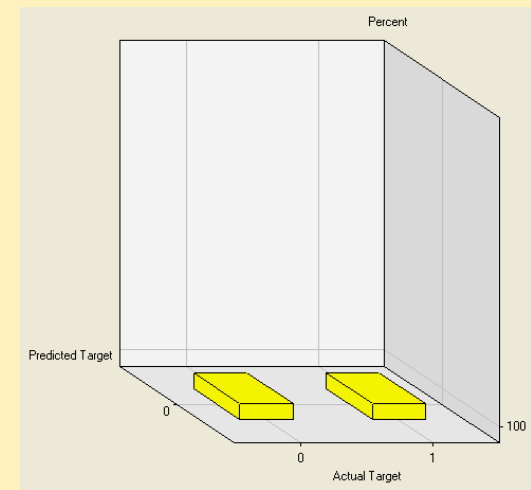
## Tree



## Regression



## Neural Net



Best



Worst

# Why SAS: 26 Years Building Relationships with Customers

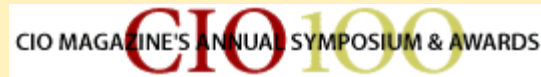


- Who?
  - 90% of Fortune 500
  - 97 of Forbes Super 100
- How many?
  - 39,500 customer sites
  - More than 3.5 million users
- Where?
  - 117 countries

# Why SAS: Award-Winning Software and Services

## Technology

## Solutions



InformationWeek 500




## Why SAS Financial Services

- More than 25 years of experience in the financial services industry.
- We work with over 2,000 Financial Institutions worldwide, including 97 percent of FORTUNE Global 500 banks.
- Recognized as the provider of the most reliable and proven analytical expertise in the world.

## What solutions we are working on

- Anti-Money Laundering
- Fraud/Risk ( $6\sigma$  control & detection)
- Credit Card Retention
- Churn Analysis
- Customer Attrition
- Credit Scoring
- Cash (ATM) Optimization

A close-up, artistic photograph of a man's face wearing glasses. The lenses of the glasses are filled with silhouettes of several people of various heights, suggesting a group or a community. The background is a soft, warm yellow gradient.

*The Power to Know™*