

A conceptual illustration of a man in a white shirt and tie, holding a glowing orb. The background is filled with various data visualizations, including a globe, a bar chart, and binary code, suggesting a focus on data analysis and technology.

# Los Modelos de Propensión a la Fuga y Desgaste de Clientes (Churn Analysis for Banking)



Viterbo H. Berberena González  
Doctor en Ciencias Técnicas  
Julio 2004

# Componentes del Problema Técnico

- Data mart analítico.
- Interfase de extracción de datos.
- Integración de las fuentes de datos externas.
- Desarrollo del modelo para medir la propensión al desgaste o la fuga 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).
- Exploración estadística de los datos.
- 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

- El riesgo constante de perder clientes.
- Los factores que más influyen están relacionados con:
  - Los productos (condiciones desfavorables o irrelevantes).
  - Las campañas (traslado, solvencia o eventos de etapas de vida).
  - Los factores externos (la competencia, las tasas de interés o la legislación).
- Las principales razones para el desgaste (attrition) son:
  - Mejores tasas de interés.
  - Mejoramiento de la condición del crédito.
  - Refinanciamiento (rollover/balance transfer).
  - Menores pagos mensuales.
  - Crédito consolidado y otros.

# Formulación del Problema de Negocio

- Pérdida de ingresos (por parte del banco) de la cuenta del producto (product attrition).
- Cierre de la cuenta del producto por parte del cliente (product churn).
- Inactividad (reducción del uso del producto, bajo cierta métrica predefinida).
- La decisión de desgaste (attrition) frecuentemente precede los eventos de cierre de cuenta o inactividad, por lo que este período de tiempo siempre debe ser tomado en cuenta.

# Definición de la Variable Objetivo

- Criterio Primario (target definition):
  - Poseedores del producto (clientes que solicitaron la cancelación de la cuenta).
  - Actividad de la cuenta del producto (la actividad cae por debajo del umbral; saldo, transacciones, etc.).
  
- Criterio Secundario (target subdefinición):
  - Titularidad de la cuenta (titulares con al menos 12 meses).
  - Historial de uso continuo (en uso continuo los 6 meses más recientes).
  - Tipo de producto (seleccionar en orden; platinum/gold/silver etc.).
  - Rentabilidad o ingresos (rentabilidad total, ingresos, importe de intereses pagados, u otros límites  $> 0$ , o cualquier otro valor positivo).

# 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 se desea retener?
- 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.

# 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.

# Composición de los Datos

- Información del cliente.
  - Personal, doméstica, demográficos, nivel socio-económico, tipo o clase de cliente.
- Transaccionalidad.
  - Titularidad, morosidad (atrasos, no pagos), pagos, intereses, saldos, comisiones, etc.
- Comportamiento/Riesgo.
  - Titularidad, morosidad (atrasos, no pagos), pagos, intereses, saldos, gastos (desembolsos).
- Servicios de valor agregado.

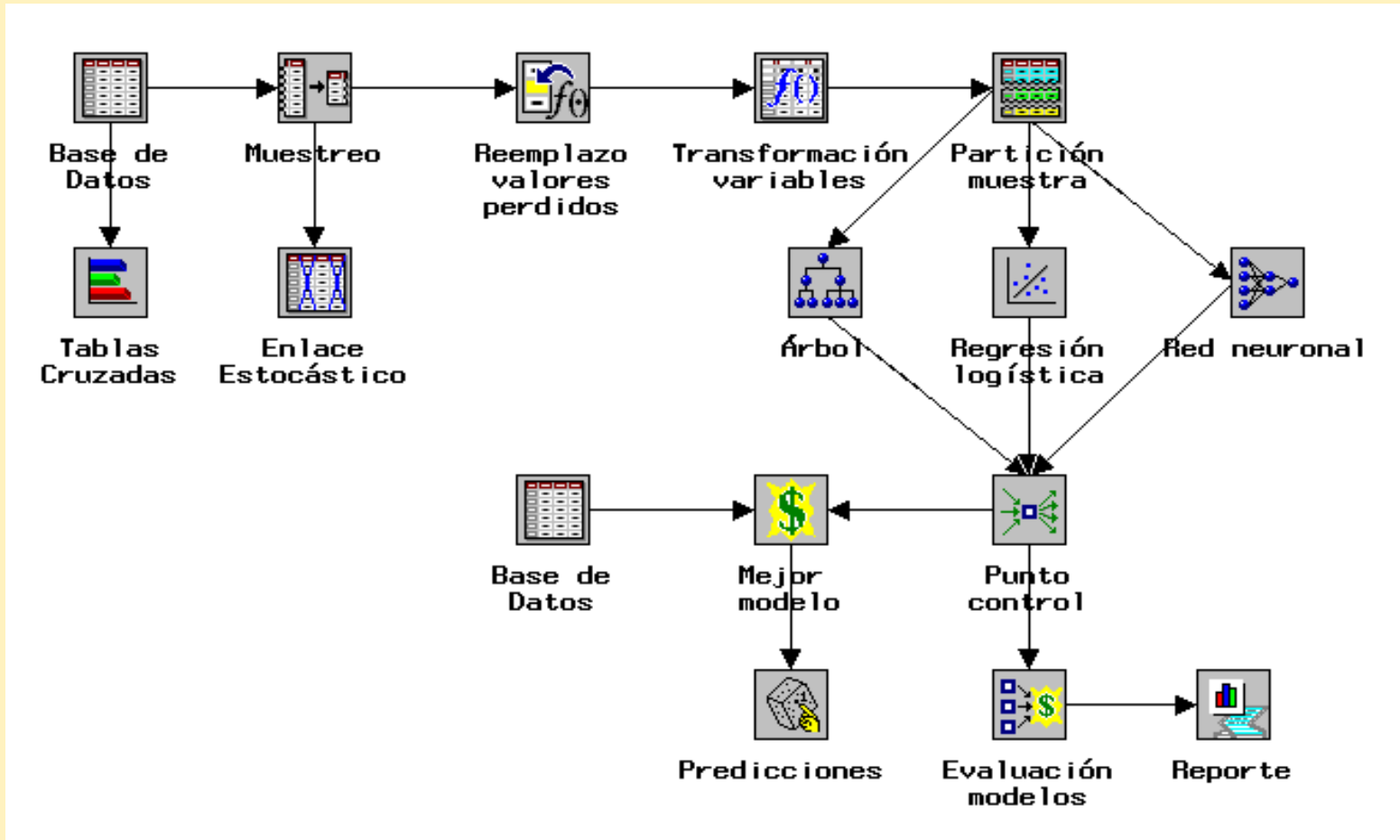
# Composición de los Datos

- Crédito.
- Desgaste/Abandono (attrition/churn).
  - Razones, titularidad, propensión, aspectos económicos, valor.
- Productos.
- Respuesta a las campañas.
  - Promociones, precios, puntos, catálogos, etc.

# 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 al desgaste/abandono (attriters, churners).
  - Creación de submodelos (si es el caso): propensión a la inactividad, baja facturación, cancelación.

# Diagrama del Flujo del Minero



# Técnicas Analíticas de Modelación

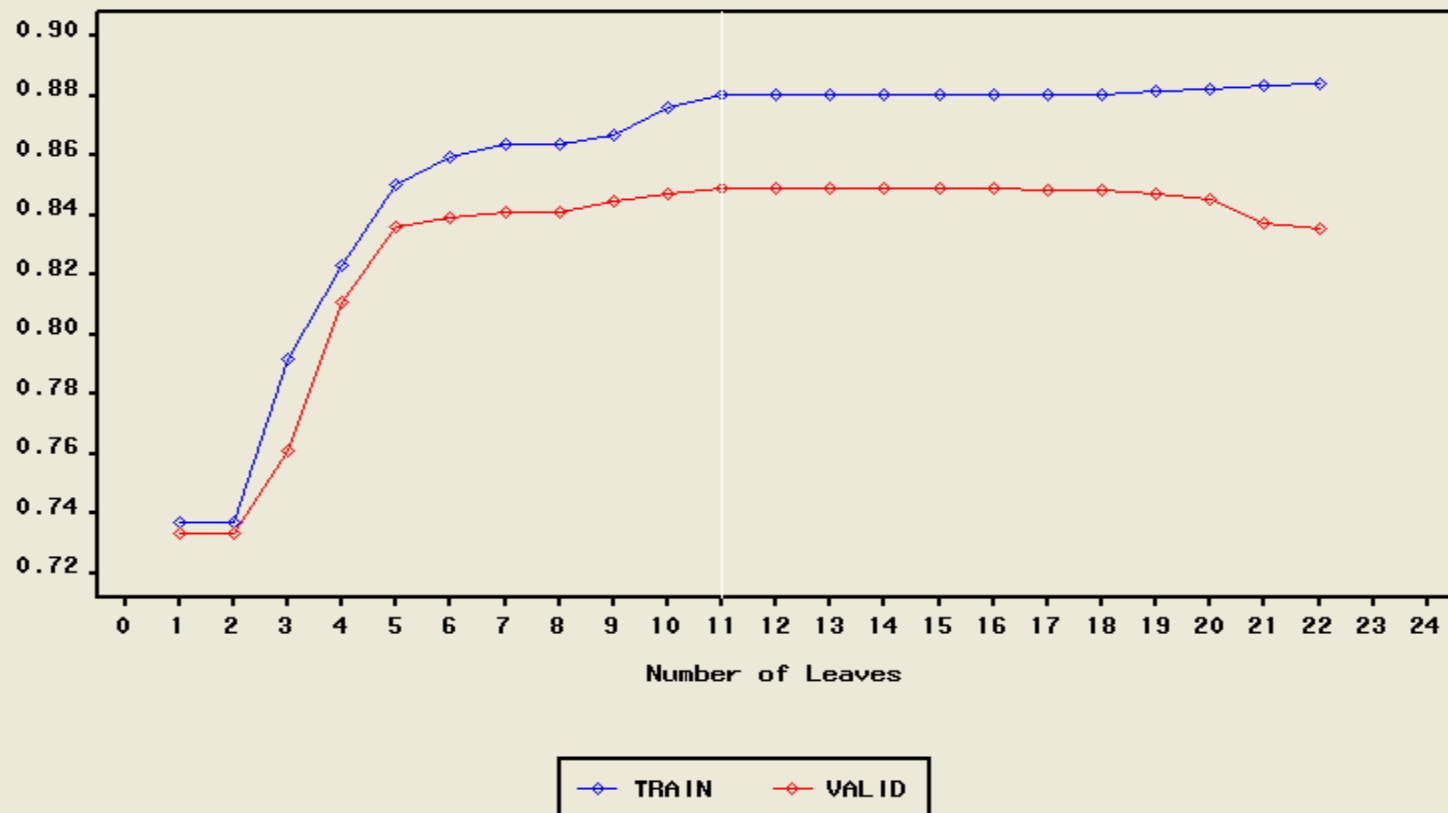
- Árboles de Decisión.
  - CHAID (chi-squared automatic interaction detection).
  - CART (classification and regression trees).
  - ML (machine learning).
- Análisis de Regresión Logística.
  - Forward.
  - Backward.
  - Stepwise.
- Redes Neuronales.
  - MPL (multilayer perceptron).
  - Ordinary RBF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

# Evaluación del Desempeño del Modelo

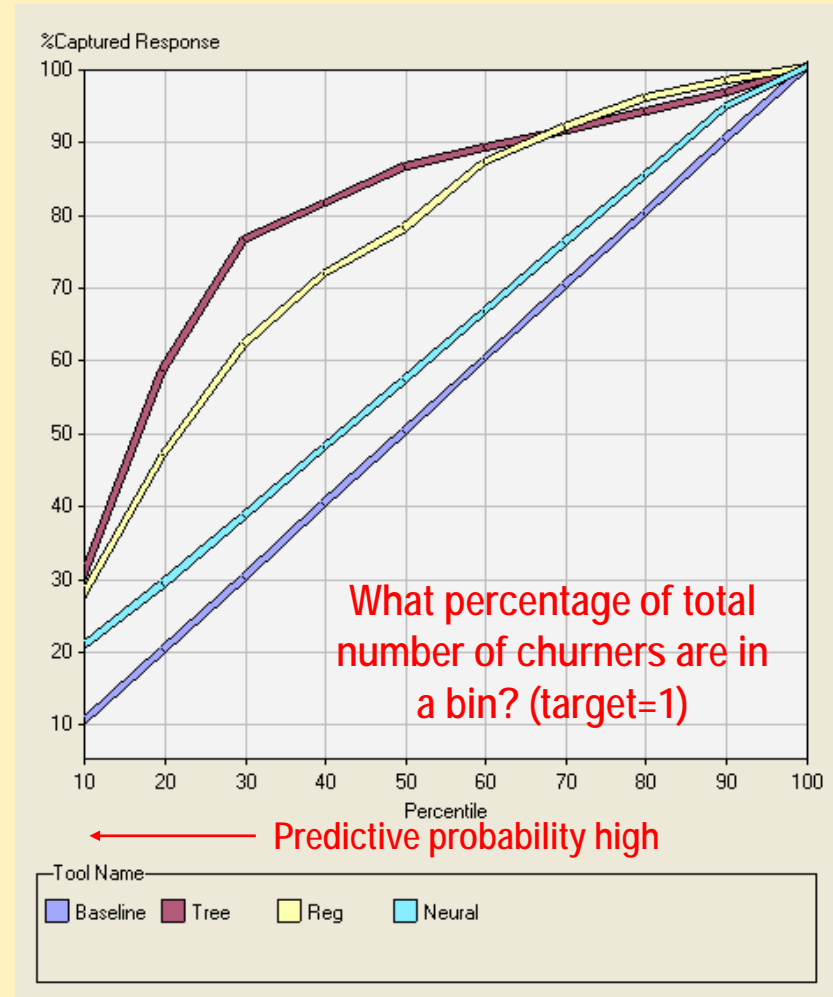
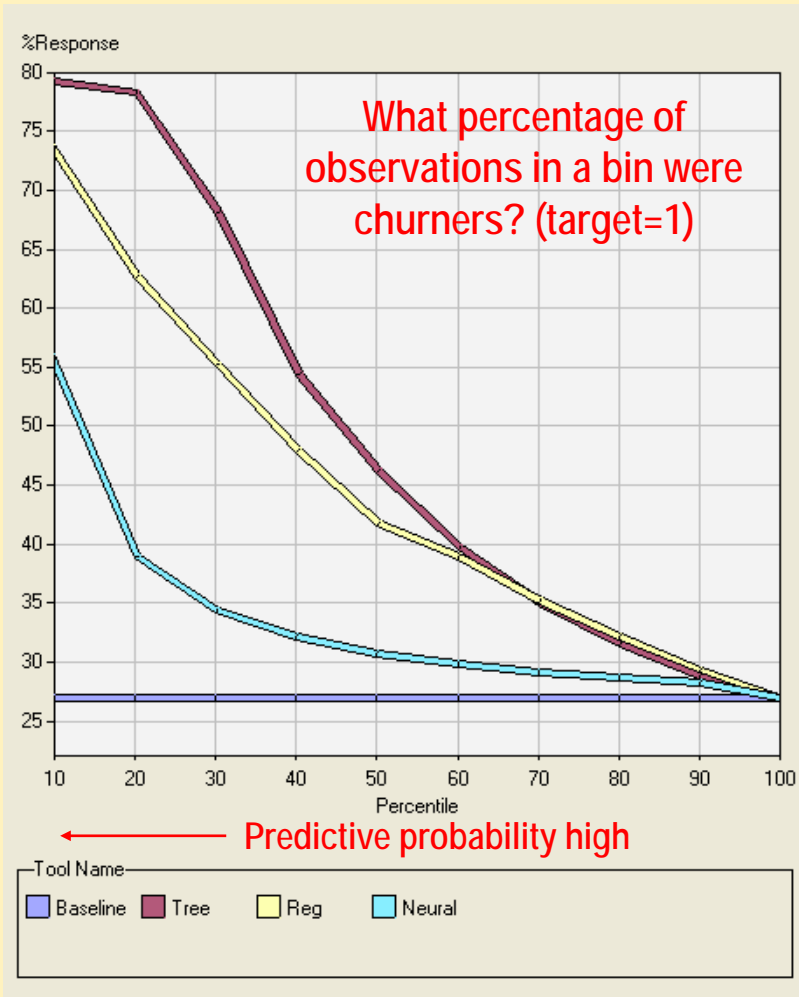
- Comparación de la capacidad predictiva de los diferentes tipos de modelos (lift chart).
- Evaluación de la calidad de las estimaciones del modelo (estimaciones de la variable objetivo).
  - Habilidad para identificar la mayor propensión (incremental lift).
  - Variables que entraron al modelo.
  - 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)

Proportion Correctly Classified



# Lift/Response Charts (cumulative)

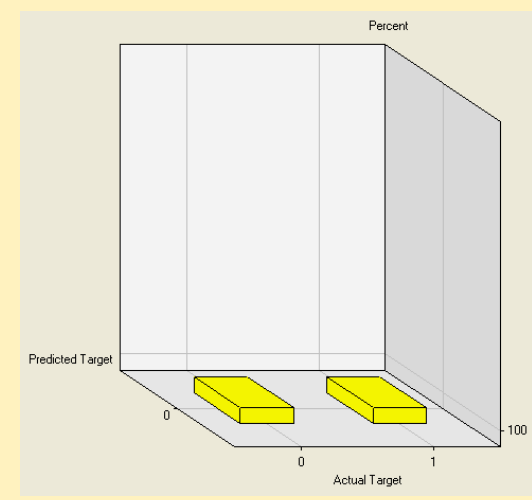
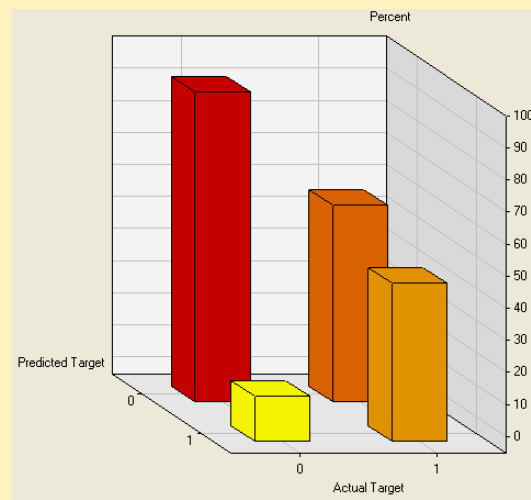
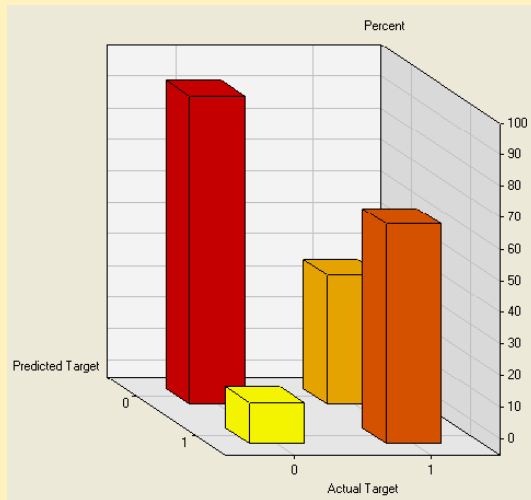


# Confusion Matrix (predictive accuracy)

**Tree**

**Regression**

**Neural Net**



**Best**



**Worst**

# Principales Resultados del Modelo

- Determinación probabilística de la propensión de los clientes a la fuga (Churn propensity).
- Pronóstico (individual) de los clientes más propensos a fugarse en la etapa.
- Mejor enfoque de las campañas de retención. De nivel grupo a nivel individual.
- Retención de los clientes valiosos para la empresa (más rentables o de mayor valor potencial).
- Aumento de la efectividad de la campañas de retención (acampañar sólo los clientes deseados).
- Elevación de la eficiencia de las campañas de retención (menores costos).

# Conclusiones

- Los modelos predictivos de minería de datos sustituyen a los métodos de pronósticos tradicionales, ya que permiten enfocar con mayor precisión el cálculo de los clientes que abandonarían el producto o servicio.
- El enfoque de minería de datos ya se está usando para la solución de problemas de fugas y campañas en retención en bancos como Capital One.
- En Banamex existe la posibilidad de aplicar estas técnicas de modelación predictiva de alta precisión y confiabilidad con el objetivo de aumentar la rentabilidad y las ganancias del banco.

# Why SAS: 26 Years Building Relationships with Customers

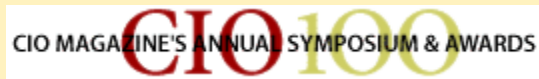


- Who?
  - 90% of Fortune 500
  - 97 of Forbes Super 100
- How many?
  - 39,500 customer sites
  - More than 3.5 million users
- Where?
  - 117 countries

# Why SAS: Award-Winning Software and Services

## Technology

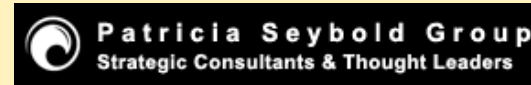
## Solutions



InformationWeek 500



**ECOMWORLD**



**Network Computing**  
COMPANIES TO WATCH



# Why SAS Financial Services

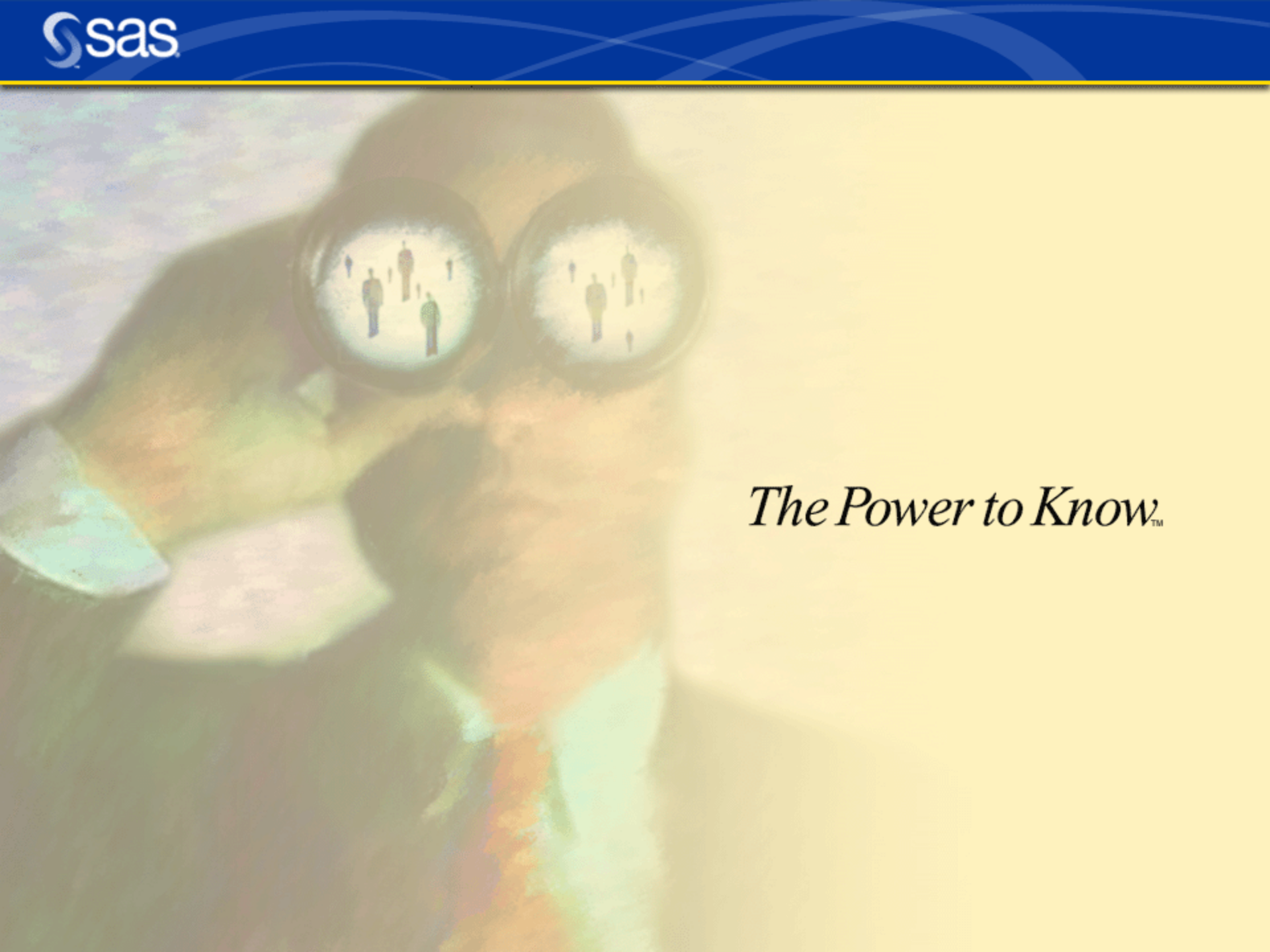
- More than 25 years of experience in the financial services industry.
- We work with over 2,000 Financial Institutions worldwide, including 97 percent of FORTUNE Global 500 banks.
- Recognized as the provider of the most reliable and proven analytical expertise in the world.

# 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

# Datos Personales

- Viterbo H. Berberena González.
- Doctor en Ciencias Técnicas.
- Director de Minería de Datos de Pearson S.A. de C.V.
- Homero 223, PH, Colonia Polanco.
- CP 11560, México D.F.
- Teléfonos:(52)(55) 5531-5324, 5531-5560 ext. 145
- Fax: (52)(55) 5203-8230
- [mailto: vberberena@pearson-research.com](mailto:vberberena@pearson-research.com)
- USA phone: (305) 390 8242 Miami.
- 1-800 711 7709 –TOLL FREE.



*The Power to Know™*