



## Minería de Datos (*Data Mining*)

# ANÁLISIS AVANZADOS DE GRANDES VOLÚMENES DE DATOS EN EL SECTOR SEGUROS (2ª PARTE)

Con esta segunda parte se finaliza el artículo de Data Mining cuya primera entrega fue publicada en el número 24 de *ACTUARIOS*, correspondiente a abril-mayo de 2006.

► **Índice:** 1. Resumen. 2. Gestión de campañas sofisticadas. 3. Estimación de provisión de siniestros pendientes. 4. Coordinación de los departamentos de marketing y actuarial. 5. Implementación de proyectos de minería de datos: a) Acceso a los datos, b) Data warehousing, c) Análisis de los datos mediante la metodología SEMMA, d) Generación de informes de resultados, e) Explotación de resultados. 6. Resumen.

► El contenido de «Minería de Datos en el Sector Seguros» ha sido proporcionado por Carmen López, Presales Consultant de Inteligencia Analítica, y por Isidoro Cortés, Insurance Industry Manager. SAS España.

### 1. Introducción

A lo largo de este segundo capítulo analizaremos la importancia de conocer a nuestro cliente, de tratarle de forma personalizada y mantener con él una relación duradera en el tiempo. Una gestión individualizada y automatizada de las comunicaciones es clave para conseguir este tratamiento individualizado. En este sentido las soluciones de gestión de campañas permiten gestionar de una forma completa y prácticamente automática todas las comunicaciones y campañas dirigidas a nuestros clientes, recogiendo su reacción ante las mismas e in-

cluyendo su feedback en la definición de la siguiente actuación comercial o de marketing.

De cara a la consecución de beneficios por parte de las compañías aseguradoras, un cálculo preciso de provisión por siniestros es estratégico. A lo largo del capítulo se detalla como las técnicas de minería de datos pueden modelizar el riesgo asumido facilitando el cálculo de esta provisión.

Para acabar introducimos el modelo SEMMA de implantación de proyectos de minería de datos, que a través de técnicas estadísticas y visuales guiará a los analistas en el muestreo, exploración, modificación, modelización y estimación de modelos de minería.

### 2. Gestión de campañas sofisticadas

Establecer una relación con el cliente es básico para el negocio en todo tipo de actividad empresarial. Las organizaciones pequeñas pueden establecer una

relación individual con cada uno de sus clientes, pero cuando la compañía crece los departamentos de marketing a menudo empiezan a pensar en términos de «desarrollo de productos» en lugar de en términos de «mantenimiento de la relación con el cliente». Es habitual que los departamentos de ventas y marketing se centren en las ventas que se consiguen con un producto en lugar de en lo bien que ese producto satisface las necesidades del cliente individual.

En un entorno como el actual, en el que los mercados están saturados y las campañas de marketing masivas son cada vez menos eficaces, la minería de datos facilita a las compañías aseguradoras la personalización de sus campañas, orientándolas hacia el cliente individual. Las estrategias de marketing basadas en el análisis de los datos de cliente están ayudando a reorientar la gestión de las relaciones con los mismos, volviendo a colocar al cliente en el centro de la estrategia de la compañía.

## Orientación al cliente

Las principales compañías aseguradoras están transformando sus antiguas estructuras orientadas a productos en estructuras centradas cliente. La minería de datos ayuda a conocer mejor al cliente identificando cuáles son sus necesidades actuales y futuras y pudiendo predecir cuáles serán sus reacciones ante ciertos cambios del mercado. El análisis del resultado de las campañas de marketing proporciona una valiosa información que se utilizará como base para el desarrollo de futuras campañas.

## Marketing: Otra frontera para la automatización

Conocer a los clientes en profundidad (sus necesidades, deseos y respuestas) es posible gracias al conocimiento que proporcionan metodologías tales como la minería de datos.

Una vez que hemos seleccionado los clientes a los que nos queremos dirigir, sabemos el producto que ofertaremos y el canal por el que nos comunicaremos con ellos, existen soluciones de gestión de campañas que permiten automatizar la gestión de nuestras actuaciones comerciales. La clave de las soluciones de gestión de campañas es que permiten controlar las respuestas de los clientes y la eficiencia de las acciones comerciales emprendidas.

La solución de marketing automation muestra cómo nuestro cliente reacciona ante las diferentes propuestas que le hacemos, a la vez que facilita la gestión de estas campañas o comunicaciones. Además se podrán tomar decisiones casi «sobre la marcha», bien para modificar la oferta o para modificar los canales de comunicación o ambos a la vez. Estas soluciones de automatización del marketing nos ayudan a maximizar el resultado de nuestros esfuerzos, ya que nos permiten concentrarnos en las actuaciones verdaderamente rentables.

## Una solución global

Las soluciones de marketing automation dan soporte al «ciclo de vida» completo de toda actividad de marketing directo: análisis, planificación, ejecución y evaluación, no sólo a una o dos fases del mismo.

Desafortunadamente, en muchas ocasiones las soluciones de gestión de campañas contienen soluciones separadas no integradas en el conjunto de la solución. Hablamos de soluciones individuales para la consulta, actualización, análisis y extracción de los datos. A menudo, estos sistemas individuales requieren considerables adaptaciones para que puedan trabajar de forma conjunta. Para evitar estos problemas, es necesario disponer de una solución de marketing automation que integre el acceso a la información, la minería de datos y la gestión de campañas proporcionando una solución única, eficaz y coherente. Si las distintas soluciones no se diseñan de forma explícita para funcionar juntas, pueden surgir problemas de portabilidad y movilidad.

## Administración eficaz de campañas

Las soluciones de integración de acceso a los datos, el Data Warehousing, la minería de datos y las soluciones de gestión de campañas permiten a los profesionales del marketing utilizar modelos de minería de datos preestablecidos dentro de ese sistema de gestión de campañas. Los responsables de marketing podrán seleccionar dichos modelos de una lista y aplicarlos a los subconjuntos de público objetivo identificados por el sistema de gestión de campañas.

Normalmente, el «código de scoring» se ejecuta sobre el subconjunto seleccionado dentro de la solución de minería de datos y el archivo final se devuelve entonces al usuario de marketing, quien, de esta manera, refina sus campañas de marketing «dirigido». A menudo se hace referencia a

esta forma de integración como «scoring dinámico», porque refleja la ejecución en tiempo real del «scoring de puntuación».

## 3. Estimación de la provisión de siniestros pendientes

Las liquidaciones de siniestros a menudo se retrasan en el tiempo. Por ejemplo, en los seguros de responsabilidad civil, el coste de un siniestro puede no conocerse hasta años después de que se haya informado del mismo. En casos de responsabilidad del empleado, hay incluso retrasos en el informe del siniestro. Dichos retrasos implican una distribución no normal de los siniestros, especialmente distribuciones distorsionadas y sesgadas según el tiempo y el tipo de negocio.

A menudo lo que se hace es utilizar una estimación de la coste del siniestro hasta que el valor real esté disponible. La estimación puede depender de lo siguiente:

- Coste del siniestro.
- Periodo medio de terminación de siniestros.
- Efecto que tendrán variables financieras como la inflación y los tipos de interés.
- Efecto que tendrán los hábitos sociales. Por ejemplo, el cambio en la percepción del hábito de fumar ha tenido un fuerte impacto en la industria del tabaco.

## Predicción coste final de los siniestros

La provisión de siniestros, necesaria para continuar con la operativa de la empresa, se fija mediante la estimación de los siniestros asegurados. La precisión de la provisión de siniestros es importante, porque normalmente los fondos destinados al pago de siniestros no pueden invertirse a largo plazo con el fin de obtener rentabilidad. Si la provisión de siniestros pendientes es demasiado baja, la empresa

puede experimentar problemas financieros. Al contrario, si la provisión es demasiado elevada, la empresa puede dejar de ser rentable. Por lo tanto, la estimación de la provisión influye en los beneficios de la compañía. El análisis de la distribución de siniestros según los clientes, la geografía y el tiempo, puede ayudar a alcanzar mejores estimaciones de esta provisión.

La minería de datos puede utilizarse para definir la distribución de los siniestros y el patrón de pagos de las prestaciones pendientes.

Primero se analizan y modelan los datos, tras lo cual se desarrolla un modelo predictivo. Después se puntuarán los siniestros pendientes en la actualidad. Concretamente, los parámetros del modelo y los datos de los siniestros se utilizan para predecir el coste final de los siniestros que todavía están pendientes. Esta estimación del coste final puede utilizarse para fijar la provisión de siniestros.

### Actualización del modelo predictivo

El cálculo de la provisión para siniestros generado mediante un modelo predictivo, se basa en la hipótesis de que el futuro se asemejará al pasado. Si el modelo no está actualizado, con el tiempo dicha hipótesis derivará en que el futuro se parecerá al pasado lejano.

A medida que se disponga de datos actuales, el modelo predictivo de minería de datos podrá actualizarse y la hipótesis pasará a ser que el futuro se asemeja a un pasado reciente.

Las técnicas de minería de datos permiten a los analistas de seguros comparar modelos y evaluarlos basándose en su rendimiento. Cuando el modelo recién actualizado supera al modelo anterior, es el momento de cambiar al nuevo modelo. Gracias a las nuevas tecnologías, los analistas pueden controlar los modelos predictivos y actualizarlos como sea necesario.

## 4. Coordinación de los departamentos de marketing y actuarial

Los departamentos de marketing y actuarial son áreas de la compañía interdependientes, las acciones de uno afectan al otro. La coordinación de esfuerzos de ambos departamentos es estratégica y tiene incluso repercusión en los ingresos y beneficios de la compañía. Aunque esta interdependencia siempre ha estado ahí no siempre se ha sido reconocida, ya que por ejemplo, tradicionalmente, el área actuarial había calculado las primas a los clientes sin contar con el apoyo del área de marketing.

La coordinación de esfuerzos se puede conseguir a través del uso de técnicas de minería de datos. El departamento de marketing puede utilizar el conocimiento generado por el departamento actuarial y los resultados de las campañas de marketing pueden convertirse a su vez en datos para investigaciones actuariales adicionales.

## 5. Implementación de proyectos de minería de datos

Se ha escrito mucho acerca del mejor modo de implementar proyectos de minería de datos. Muchos libros y trabajos escritos por expertos en IT tratan el tema en detalle. Una idea común en la mayoría de autores es que en un proyecto de minería de datos es imprescindible no perder de vista la situación real del mercado; los modelos generados deben ser flexibles y deben poder ser actualizados con el tiempo. Además debe tenerse en cuenta el «desafío práctico» que implica implementar un proyecto de este tipo. Un enfoque centrado en los datos es especialmente eficaz. Podemos hablar de las siguientes fases:

- a) Acceso a los datos.
- b) Data Warehousing.
- c) Análisis de los datos.

- d) Generación de informes donde se incluya el resultado de los análisis.
- e) Explotación de resultados para obtener ventajas competitivas.

### a) Acceso a los datos

Partir de una información fiable y precisa es imprescindible en todo proyecto de minería de datos. Una solución completa de acceso a los datos debería incluir los siguientes elementos:

- Acceso a todos los tipos de fuentes de datos.
- Acceso a todas las fuentes de datos independientemente de la plataforma en la que residan.
- Preservación de la fuente de datos mediante el uso de rutinas de seguridad.
- Una GUI coherente y fácil de usar que, sin requerir un conocimiento extensivo de cada tipo de datos, proporcione la flexibilidad necesaria para satisfacer necesidades específicas.
- Integración con la tecnología existente en lugar de rutinas de acceso que requieren cambiar las soluciones de hardware y software o un aprendizaje adicional por parte de los usuarios.

Un almacén de datos adecuadamente diseñado e implementado puede ayudar a satisfacer estas premisas básicas de toda estrategia de acceso a los datos.

### b) Data Warehousing

Disponer de un Data Warehouse permite a los investigadores acceder fácilmente a los datos, que podrán estar almacenados en diferentes tablas alojadas en distintas plataformas. El Data Warehouse permite a los analistas combinar y agregar datos a distintas áreas temáticas. No obstante, antes del análisis, hay que depurar (o en caso extremo eliminar) aquellos que contienen errores, valores ausentes... Sin embargo, los resultados pueden ser tendenciosos puesto que los datos

eliminados podrían haber formado parte de relaciones importantes.

Las principales tareas de depuración, como hacer que haya coherencia entre los nombres de las variables, atribuir valores ausentes, identificar errores, corregirlos y detectar valores extremos, pueden realizarse con relativa facilidad utilizando la minería de datos.

Una adecuada solución de Data Warehousing puede mejorar considerablemente el resultado de todo el proyecto de minería de datos, ya que se podrán obtener mejores resultados con mayor rapidez y, a menudo, a un coste más bajo.

La Solución de Data Warehouse de SAS proporciona un entorno visual para la gestión de almacenes de datos. Mediante el uso de esta solución, los analistas especifican los metadatos que definen fuentes y almacenes de datos, bibliotecas de código y otros recursos de Data Warehouse. A continuación, la Solución ETL de SAS utiliza estos metadatos para generar o recuperar código, que extrae, transforma y carga los datos en el Data Warehouse y en los «data marts».

### c) Análisis de datos mediante la metodología SEMMA

Incluso después de que se hayan combinado los datos y se hayan agregado a áreas temáticas, ver las relaciones muy complejas sigue siendo difícil. No basta con inspeccionarlos visualmente ni con aplicar análisis estadísticos simples. Patrones reales pueden parecer aleatorios y no detectarse. Además, los modelos lineales difícilmente pueden describir relaciones no lineales. Las técnicas de minería de datos pueden superar estas limitaciones empleando los siguientes métodos:

- Herramientas de trazado y exploración de datos de sofisticadas GUI (interfaz gráfica de usuario) para mostrar mejor las relaciones entre variables.

- Metodologías de selección de variables con las que identificar las variables más importantes que deben ser incluidas en los modelos.
- Técnicas de modelización avanzada tales como modelos lineales con interacciones.
- Redes neuronales no lineales y modelos de árboles.
- Técnicas de evaluación que ayudarán a los analistas a seleccionar el modelo más rentable basándose en criterios de pérdidas y beneficios.

Una vez que se ha accedido a los datos, éstos pueden explorarse utilizando GUIs que emplean sofisticados algoritmos de minería de datos. Por ejemplo, los subconjuntos de datos pueden revelar importantes relaciones que podrán ser utilizadas en la definición de las campañas de marketing. Al desagruparlos por región y por empresa se pueden descubrir anomalías en las operaciones que implican incremento en los costes. Además profundizar en los datos puede ayudar a descubrir oportunidades de obtención de beneficios desconocidas hasta el momento.

Dentro de un proyecto de minería se incluye además del análisis de datos, la selección, exploración y modelización de los mismos. A través de este análisis se puede descubrir información oculta que podrá ser utilizada para obtener ventajas competitivas.

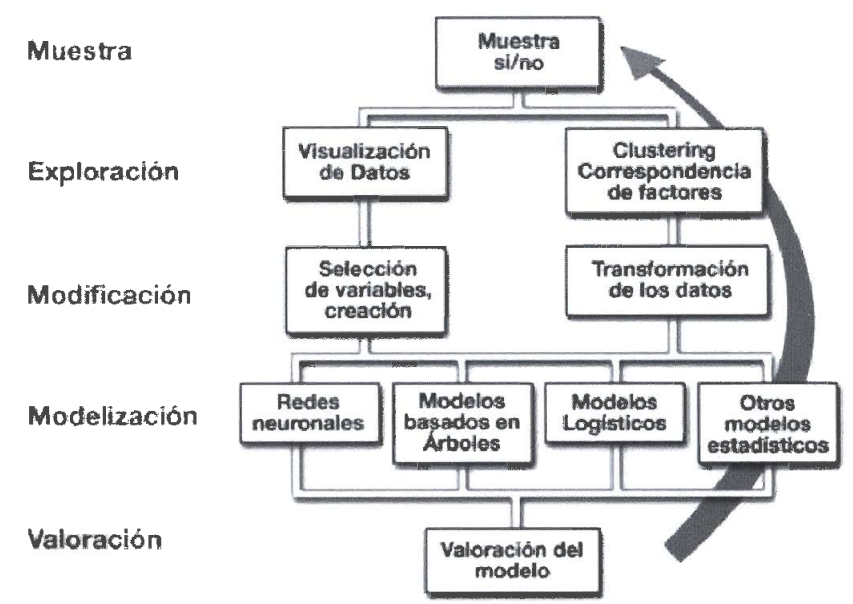
Con el fin de establecer un camino predecible y sencillo que ayude a los responsables de la minería de datos, SAS ha desarrollado un ciclo de análisis de minería de datos conocido por el acrónimo SEMMA. Este acrónimo procede de los cinco pasos de la fase de análisis dentro de un proyecto de minería de datos. Estos cinco pasos son:

- Muestreo.
- Exploración.
- Modificación.
- Modelización.
- Estimación.

La imagen 1 ilustra las tareas que se incluyen en un proyecto de minería de datos y asigna dichas cinco tareas a los cinco pasos de la metodología SEMMA.

El ciclo de análisis SEMMA utiliza técnicas estadísticas y visuales pa-

## El ciclo de análisis SEMMA



ra guiar a los analistas a través del proceso de exploración de los datos. El ciclo incluye la transformación para descubrir las variables con mayor capacidad predictiva, modelando estas variables para predecir resultados y generando modelos finales. Además la metodología de análisis SEMMA también incluye la comprobación del modelo utilizando nuevos datos. El ciclo de análisis SEMMA es una ampliación moderna del método científico.

### **Muestreo**

El primer paso en la metodología es crear una o más tablas utilizando muestras de los datos contenidos en el Data Warehouse. Las muestras deberían ser lo suficientemente grandes como para contener información significativa, aunque lo suficientemente pequeñas como para poder procesarse con rapidez. Este enfoque permite obtener resultados «coste-eficientes». Al explorar una muestra representativa en lugar del volumen completo de información se reduce de forma drástica el tiempo de procesamiento requerido.

Si existen patrones generales en el conjunto de la muestra, éstos también estarán presentes en una muestra representativa. Si un nicho es tan pequeño como para no estar representado en la muestra y pese a ello tan importante como para influir en la imagen completa, podrá descubrirse utilizando métodos de sumariación.

### **Exploración**

Después de obtener una muestra de los datos, el siguiente paso es explorarlos visual o numéricamente con el fin de observar tendencias inherentes o agrupaciones. La exploración ayuda a descubrir estas agrupaciones. Si la exploración visual no revela tendencias claras, los analistas pueden explorar los datos me-

dante técnicas estadísticas que incluyen análisis de factores, análisis de correspondencia y clustering. Por ejemplo, los padres primerizos son a menudo más conscientes de la necesidad de un seguro de vida, pero puede que busquen invertir la mínima cantidad de dinero posible. Es más probable que este grupo responda a los mailings directos para un seguro de vida.

### **Modificación**

Por modificación de los datos entendemos la creación, selección y transformación de una o más variables para centrar el proceso de selección de modelos en una dirección particular o para aumentar los datos para obtener claridad o coherencia.

Basándose en los descubrimientos de la fase de exploración, los analistas pueden necesitar tratar los datos para incluir información tal como la agrupación de los clientes y subgrupos significativos o introducir nuevas variables tales como un ratio obtenido comparando dos variables anteriormente definidas. Puede que los analistas también necesiten buscar valores extremos y reducir el número de variables para limitarlas a las más significativas. Además, dado que la minería de datos es un proceso dinámico e interactivo, a menudo es necesario modificar los datos cuando los datos anteriormente extraídos sufran algún cambio.

### **Modelo**

Crear un modelo de datos implica la utilización de una solución de minería que busque automáticamente una combinación de datos que prevean de forma fiable un resultado deseado.


Después de que se haya accedido a los datos y éstos se hayan modificado, los analistas pueden utilizar técnicas de modelización para construir modelos que expliquen patrones.

Las técnicas de modelización en la minería de datos incluyen redes neuronales, modelos basados en árboles, modelos logísticos y otros modelos estadísticos tales como análisis de series temporales y análisis de supervivencia.

### **Estimación**

El siguiente paso en todo proyecto de minería de datos consiste en estimar el modelo para su posterior evaluación. Un método común para evaluar un modelo es aplicarlo a la porción de los datos que se dejaron de lado durante la etapa de muestreo. Si el modelo es válido debería funcionar para esta muestra reservada, de igual modo que funciona para la muestra utilizada para construir el modelo.

De forma similar, los analistas pueden probar el modelo utilizando datos conocidos. Por ejemplo, si se sabe qué clientes tuvieron altos índices de retención y el modelo se ha definido para predecir la retención, los analistas pueden comprobar si el modelo señala a esos clientes. Además, la aplicación práctica del modelo, por ejemplo hacer mailings parciales en una campaña de marketing directo, ayudan a probar su validez.



***La clave para obtener una ventaja competitiva en el sector seguros reside en asumir que las bases de datos de clientes si se gestionan, analizan y explotan de la forma adecuada, son bienes únicos y valiosos.***

## Iteración

Aunque estimar los modelos de datos es el último paso en la metodología SEMMA, estimar la eficacia de modelos de datos a menudo no es el paso final en una implementación real de SEMMA.

Como SEMMA es un ciclo, los pasos internos se suelen realizar iterativamente dentro del conjunto del proyecto de minería de datos.

### d) Generación de informes de resultados

La capacidad de generación de informes de una compañía es esencial para dar utilidad a los datos, permitiendo a los usuarios crear y publicar informes directamente desde los almacenes de datos y otras fuentes de información.

Las características clave de un sistema de generación de informes deberían incluir:

- Completa integración con el sistema.
- Facilidad de uso mediante interfaces gráficas de usuario.

- Distribución rápida de informes.
- Compatibilidad con la Web.

Una solución de generación de informes completa, como la incluida en SAS® Enterprise Miner™, permite a los usuarios de negocio crear, publicar e imprimir informes en un formato enriquecido partiendo de la información guardada en sus Data WareHouses. Mediante interfaces fáciles de usar, los usuarios tienen la habilidad de crear gráficos, tablas, diagramas y texto dentro de un único informe y desde sus propios PC's.

### e) Explotación de resultados para obtener ventajas de negocio

La información obtenida a partir de la minería de datos puede incorporarse en un informe ejecutivo o procesamiento analítico en línea y, a continuación, difundirse por toda la organización. La dirección de la compañía puede utilizar la minería de datos para realizar los análisis clave para la planificación y acción estratégica, ¿cómo podemos aumentar el retorno del capital invertido en nuestras campañas de marketing? La

minería de datos ayudará a las aseguradoras a estar más preparadas para su crecimiento futuro.

## 6. Resumen

La clave para obtener una ventaja competitiva en el sector seguros reside en asumir que las bases de datos de cliente si se gestionan, analizan y explotan de la forma adecuada, son bienes únicos y valiosos.

La minería de datos puede ayudar a las compañías de seguros a acceder a la inteligencia contenida en sus bases de datos de cliente.

La minería segmenta la base de datos, genera modelos predictivos, analiza el mercado y proporciona a la compañía el conocimiento necesario para responder con rapidez y eficacia preguntas estratégicas de negocio. A partir de la información generada se pueden desarrollar nuevos productos e implementar las estrategias de marketing más adecuadas permitiendo transformar «riqueza de información» en «riqueza de predicción, estabilidad y beneficios». ●

Consulta [www.actuarios.org](http://www.actuarios.org)  
nuestra  
página  
web

actuari@s  
en la red

