

# Los Modelos de Minería de Datos en la Industria de los Medios de Comunicación

Viterbo H. Berberena  
Pearson

**L**a industria televisiva en México ha sufrido grandes cambios en los últimos años. Dentro de este proceso se puede mencionar el incremento del nivel de competencia que han mantenido las empresas líderes en el mercado. Como resultado de esta situación, las televisoras se han dado a la tarea de buscar nuevas formas de promocionar sus servicios a las compañías necesitadas de publicidad masiva.

Dicen algunos investigadores de los medios que «el espectador es lo que ve», refiriéndose a que los hábitos de los televidentes tienden a revelar sus características personales, gustos y deseos.

Los patrones del espectador incluyen características tales como: tipo de programa, frecuencia de observación, la hora de inicio del programa, la duración del programa, etcétera. Las técnicas de minería de datos permiten identificar conjuntos de televidentes que son el objetivo de los anunciantes y las cadenas televisivas.

La minería de datos es una nueva clase de métodos analíticos inteligentes que tienen la capacidad de extraer patrones de comportamientos muy complejos, ocultos en montañas de datos. Estos datos, a que hacemos referencia, se obtienen, por ejemplo, a partir de las mediciones de audiencias de las televisoras, así como de encuestas y otros instrumentos de medición disponibles. En nuestro trabajo académico en el área de Modelos Decisionales de Mercadotecnia hemos encontrado que las televisoras poseen una gran cantidad de datos históricos de los televidentes que no se aprovechan a plenitud, de hecho hemos realizado proyectos de extracción de patrones (sin fines comerciales, sólo académicos) con excelentes resultados, partiendo de que podemos determinar con muy alta precisión el rating de nuevos programas lo que permite a las televisoras determinar la audiencia objetivo con gran exactitud y fijar adecuadamente los precios de la publicidad asociada.

Con este artículo no se pretende hacer una disertación detallada del uso de los modelos de minería de datos en la Industria de los Medios de Comunicación. Sólo se intentará introducir al lector al mundo de los

modelos de soporte de minería de datos, basados en la investigación de patrones de bases de datos.

## Los Patrones de Consumo del Espectador

La nueva generación del mercado masivo de los medios, compuesta por millones de consumidores, tiene que ser diseñada para registrar, en modelos, los hábitos de observación y consumo del espectador. Los perfiles de los teleespectadores, que contienen la edad, los ingresos, la condición familiar, datos psicográficos y demográficos, deben ser calculados y puestos a disposición de los vendedores, publicistas, programadores y demás personas involucradas en la industria de la multimedia. Llevamos más de cinco años trabajando en el desarrollo de modelos de minería de datos para diferentes empresas y organizaciones. Por razones obvias, nuestra experiencia acumulada y el aval de nuestros clientes de minería de datos no son tema para este artículo, que es más bien informativo que técnico. Además, desde el punto de vista académico nosotros coordinamos la Especialidad Ejecutiva en Minería de Datos en el Centro de Alta Dirección en Ingeniería y Tecnología de la Universidad Anáhuac en conjunto con el SAS Institute, donde se puede adquirir una preparación básica a nivel ejecutivo.

La televisión que conocemos hoy experimenta una transformación importante, permitiendo aprovechar el poder del mercadeo directo «uno a uno», asociado con los medios en línea.

La industria de los medios examina con gran atención los proyectos tecnológicos conjuntamente con los planes de negocio, usando información proveniente de los sitios de Internet de las compañías de Mercadeo de Televisión Interactiva, los sitios de los socios tecnológicos, los reportes de las compañías de investigación de mercados y la información de la industria publicitaria. A partir de esta gran «explosión» de datos disponibles, las empresas multimedia tienen que utilizar sistemas de plataformas operacionales tipo ERP (Enterprise Resources Planning) para poder operar adecuadamente. Estas plataformas generan grandes bases de da-

tos que pueden ser utilizadas para la extracción de reglas, patrones, regularidades, relaciones y restricciones necesaria para la toma de decisiones y ocultas a las técnicas convencionales de análisis.

La televisión multicanales de hoy en día se está transformando a través de la computarización y sofisticados paquetes de tecnología digital con el objetivo de recopilar información de los consumidores individuales y sus familias. Esta información se convertirá, en el mejor de los casos, en perfiles de consumidores que se usarán para enfocar al mercado, de forma que se pueda personalizar la publicidad. Esto ya se realiza en Europa y los Estados Unidos. Sin embargo, en México también se puede hacer, puesto que las televisoras cuentan con una inmensa cantidad de datos de muy buena calidad que se pueden utilizar para extraer patrones de comportamiento, aunado con la capacidad que hay en el país para realizar este tipo de investigación analítica del más alto nivel.

La misma tecnología que se usó inicialmente en Internet para la detección de intrusos, que incluye la minería de datos, la modelación de los usuarios y el uso de agentes electrónicos inteligentes, ha sido adoptada por la industria de la televisión en los países del primer mundo para la planeación, diseño y selección de la programación. Esta nueva infraestructura de recolección de datos permitirá lo que la industria de la televisión está llamando «T-Commerce» (for Television Commerce). Las compañías involucradas en estas nuevas técnicas de recolección de datos incluyen algunas de las más conocidas corporaciones de los Estados Unidos, como Microsoft, AT&T, Liberty Media, Procter and Gamble, NDS (Rupert Murdoch), Cisco, A.C. Nielson, Scientific-Atlanta, and Young and Rubicam. Los modelos que estas compañías desarrollan combinan los aspectos más valiosos de la Internet y los medios masivos, así como los nuevos sistemas que han sido diseñados para monitorear, no sólo cada actividad de los usuarios y su navegación en la red, así como los programas y los comerciales que ven también. Como consecuencia, esto genera de nuevo un cuestionamiento sobre la privacidad de los individuos.

Para los anunciantes, el desarrollo de una tecnología que combina la interactividad de los sitios de Internet con los elementos de la televisión dedicados a la publicidad, es un sueño hecho realidad. Si revisamos los últimos trabajos de investigadores destacados en esta área observamos que la Internet y la Televisión están convergiendo hacia un concepto denominado «ITV» (Interactive Television). Existen instituciones que obser-

van de cerca estos cambios como el «Center for Digital Democracy» y otros. Las plataformas para la recopilación de datos, proveen a los anunciantes la capacidad de enfocar cada espectador individual con anuncios personalizados que incrementan la probabilidad de impulsar las compras de un determinado producto o servicio, tanto en TV como en Internet o la mezcla de ambas formas (ITV o televisión interactiva).

Cada programa sintonizado, cada anuncio observado, cada clic y cada material bajado de la red de Internet, se convierte en el medio para la recopilación de datos que servirán para la creación del perfil de los consumidores, que a su vez, marcaran el mercado objetivo de los anuncios individualizados.

A las grandes compañías multimedia no les queda de otra: «Tienen que adquirir, desarrollar y explotar una infraestructura de recolección de datos para la obtención de modelos, que le permitan, entre otras cosas, individualizar los anuncios publicitarios de productos o servicios».

Pero, ¿Cómo convertimos los datos en modelos que suministren información estratégica para la toma de decisiones en el negocio?

## **Los Modelos de Minería de Datos**

Investigaciones de este tipo han permitido a los países de primer mundo establecer técnicas de filtrado de decisiones teóricas y colaborativas en el correo directo, el monitoreo de anuncios en sitios en Internet y la construcción de guías de programas personalizados, entre otras muchas aplicaciones importantes.

Las televisoras tienen que desarrollar e implementar un sistema para la obtención de los perfiles de los espectadores que incluye, de forma básica, la capacidad analítica para la construcción de los perfiles a partir de la investigación de los patrones en las bases de datos, la posibilidad de determinar el mercado objetivo (individuos con mayor probabilidad de responder ante ciertos anuncios comerciales) y modelos para la toma de decisiones estratégicas en mercados clasificados como potencialmente divergentes.

En principio, todas las fuentes de información, tanto las de origen cualitativo como cuantitativo son válidas. Sin embargo, la minería de datos se focaliza en la extracción de patrones de comportamiento redundantes en los datos, pero no observables con métodos tradicionales de análisis, por esta razón las mediciones de tipo cuantitativo son las mejores. En los datos está el



comportamiento del usuario del producto o servicio, es decir, lo que pasó en la realidad. En contraste, otros instrumentos de recolección de datos como encuestas, por ejemplo, tienen un origen estocástico o cualitativo que reflejan más los síntomas del comportamiento que el comportamiento en sí. Otra cuestión es que las encuestas trabajan con muestras, en la mayoría de los casos, muy pequeñas, para encontrar patrones redundantes de comportamiento con herramientas de minería de datos. Acerca de los datos de la medición de audiencia que hace IBOPE, son una gran fuente de información con la que ya hemos trabajado en la predicción de ratings a nivel académico.

La minería de datos utiliza una serie de técnicas de clasificación supervisada y no supervisada para analizar los datos de los espectadores y segmentar los individuos y sus hogares. Cada clasificación constituye un perfil formado por datos demográficos (edad, género), geográficos (áreas geográficas, tipos de mercados) y psicográficos (estilos de vida, intereses personales, familiares, etcétera).

Las técnicas de minería de datos han hecho su aparición para quedarse. La urgente necesidad de las televisoras por enfocar sus mercados, desarrollar los perfiles de los telespectadores y sus familias y establecer precios adecuados según el «rating» de los programas, implican que estas compañías tienen que desarrollar sus modelos de soporte para la toma de decisiones estratégicas que le permitan estar un paso adelante de su competidor más cercano.

En el pasado, el problema de la minería de datos, era su costo. Aunque existían herramientas aparentemente útiles, las compañías no podían afrontar esos gastos. Además, el hecho de que las herramientas de minería emplean algoritmos matemáticos muy complejos, creaba una dificultad para operar con los usuarios de negocios, no técnicos. Sin embargo, esto ha cambiado y los resultados de los modelos de minería de datos son comprensibles, tanto para los técnicos en la materia, como para los especialistas de negocios que deben operar con ellos.

Un ejemplo sencillo sería la determinación estocástica de la propensión de los usuarios a responder a determinados anuncios comerciales. Como es lógico para entrenar el modelo, que pudiera ser de tipo topológico, una regresión o una red neuronal, se utiliza una base de datos que contiene un identificador de usuario, variables de clasificación y de negocios y además la variable objetivo, que dice quién (en el pasado) res-

pondió a un cierto comercial. Un criterio de calidad del modelo sería lo que se conoce como matriz de confusión, la cual nos indica que si el modelo es bueno, los usuarios que respondieron positivamente al anuncio en la realidad, reciben una alta probabilidad de responder. Por el contrario, si el modelo es malo, no es capaz de hacer una discriminación entre los que respondieron y los que no respondieron. Al final el modelo se utiliza para evaluar la propensión de respuesta de otros usuarios, a los que se les midió las mismas características que las que se usaron para entrenar el modelo. Se genera una tabla donde a cada individuo se le predice una probabilidad de responder al anuncio en cuestión.

## Conclusiones

El desarrollo de modelos de minería de datos de alta precisión y confiabilidad es una necesidad de las empresas de medios de comunicación para poder enfrentar la alta competencia de mercados cada vez más globalizados.

En México se pueden desarrollar proyectos de minería de datos para la determinación de perfiles de la audiencia y la designación de mercados objetivos para los anunciantes, de forma que se pueda optimizar la eficacia de la publicidad.

A los proyectos de minería de datos, en cualquier caso, se le puede determinar un retorno, lo que pone en evidencia la factibilidad económica de los mismos.

Como parte de nuestra colaboración académica con los departamentos de posgrado de universidades privadas importantes del país, se han realizado algunos ejercicios como la estimación del «rating» de nuevos programas a partir de patrones existentes en bases de datos. El objetivo de estos proyectos académicos fue la obtención de información que permita la elaboración de estrategias de planeación, diseño y selección de programas más rentables para estas compañías. Desgraciadamente, por cuestiones de confidencialidad relacionada con la fuerte competencia entre las empresas de medios no podemos dar detalles específicos de los resultados del trabajo. Sin embargo, podemos afirmar con confianza que, con una aplicación inteligente de Minería de Datos sobre «ratings» históricos, se podrá predecir, con bastante exactitud, el posible «rating» de un futuro programa que aún no sale al aire y adecuar los contenidos y precios de la publicidad a la audiencia objetivo.