

1 Introducción

Esta tesis tiene como objetivo mostrar la aplicación y la enorme utilidad de un conjunto de modelos de evaluación de sistemas diagnósticos con raíces en la Psicofísica y con una rápida expansión en la Medicina, en un campo en el que son relativamente desconocidas, como es el del Marketing Relacional.

El problema de evaluar la precisión de las herramientas diagnósticas (de cualquier tipo, desde una simple prueba a todo un protocolo de evaluación) es especialmente relevante en Medicina. Muy a menudo, el objetivo del sistema de diagnóstico es discriminar entre dos estados mutuamente exclusivos, tales como la presencia o ausencia de una enfermedad. Por ejemplo, decidir si en una placa de rayos X se observa un tumor o no, o por parte de un psicólogo decidir si se diagnostica a un paciente como enfermo de depresión o pánico.

Estas técnicas son conocidas como "análisis ROC" o "análisis de curvas ROC" y han permitido en Medicina y en muchas otras áreas clínicas el desarrollo de pruebas de detección rápida, o de "*screening*", puesto que permiten un enfoque común de evaluación de la capacidad diagnóstica y por tanto un desarrollo iterativo e incremental de las mismas. El empuje de esta aplicación se ha visto reflejado en el desarrollo de procedimientos prácticos, no sólo de la estimación de dichas curvas, sino de los contrastes estadísticos que permiten una interpretación similar a la de muchas otras pruebas estadísticas.

ROC es el acrónimo del inglés *Relative (o también Receiver) Operating Characteristic*. Una curva ROC es la representación gráfica de la tasa de éxito (probabilidad de que el sistema detecte una señal cuando efectivamente está presente) frente a la tasa de falsa alarma (probabilidad de decir que el sistema la detecte cuando en realidad está ausente) para detecciones sí / no de una señal que puede estar presente o ausente.

En Psicología, las curvas ROC se conocen exhaustivamente en Psicofísica, en donde el evento bajo estudio puede ser una luz o un sonido, que está ausente o presente en cada ensayo de un experimento (Green y Swets, 1966; Swets, 1973). El observador dice sí o no. La probabilidad de decir sí se puede manipular de múltiples maneras, por ejemplo, variando los pagos por los aciertos frente a los rechazos correctos.

Las curvas ROC se han utilizado en varios campos del diagnóstico médico, por su validez para comparar tanto la precisión diagnóstica de diferentes pruebas simples de laboratorio como conjuntos completos de pruebas, protocolos, etc., incluyendo datos objetivos y subjetivos.

Por otro lado, tenemos algunos de los problemas más típicos del Marketing Relacional, uno de cuyos principios básicos es mantener la fidelidad de los mejores clientes, puesto que será más beneficioso mantener a los mejores clientes, aunque cueste dinero hacerlo, que ganarle clientes a la competencia. La forma de hacerlo es manteniendo, cultivando la "relación" con el cliente, y de aquí han surgido los conceptos de CRM, "*Customer Relationship Management*". Uno de los problemas que existen dentro de este enfoque es detectar qué clientes serán contactados, qué clientes están en riesgo de abandonar su relación con nuestra empresa, en general, problemas similares conceptualmente (aunque muy diferentes en la práctica) a los que aborda la Medicina.

Un capítulo se dedica en exclusiva a este enfoque tan importante estos días en el Marketing. Se plantean muchos problemas aplicados en este campo, por supuesto no todos los cuales se podrán abordar en este trabajo. El interés por el CRM se debe principalmente a que se ha comprobado en múltiples sectores que los costes para adquirir un nuevo cliente son mucho mayores que los necesarios para mantenerlo. El objetivo es por tanto ser capaces de predecir tanto la fidelidad como el abandono, así como de elegir los factores o predictores que se puedan utilizar en la práctica

empresarial. De entre las distintas posibilidades elegiremos un enfoque fuertemente metodológico y aplicado, pero forzosamente limitado al análisis de la base de datos disponible.

Hemos de decir que el sector del Marketing, desde el punto de vista práctico, es completamente diferente al de la Medicina clínica. Es un campo de batalla empresarial en el que consultores, ejecutivos, múltiples departamentos empresariales tienen mucho que jugar. No resultará extraño entonces la enorme variedad de propuestas en todos los sentidos, los rápidos cambios de enfoque global, y lo que suelo llamar "sopa de letras", por la costumbre anglosajona de denominarlo todo con siglas, obligando a tener un glosario a modo de mapa para poder entender de qué se está hablando. Tenemos así CRM (*Customer Relationship Management*), DM (*Data Mining*), KDD (*Knowledge Discovery in Databases*).

El campo de la Medicina tiene por supuesto un ritmo diferente. Puede resultar sorprendente que el objetivo sea similar al de Marketing. Pero en ambos casos resulta imprescindible detectar: en el campo médico descartar, o sospechar, con rapidez la presencia de enfermedad; en el campo del Marketing, detectar si un cliente está próximo a abandonar.

El problema específico que vamos a abordar es el de identificar los clientes de una tarjeta de fidelización multipatrocinador con implantación en toda España, que es "Travel Club", que pueden no estar aprovechando todas las posibilidades del programa. No sabremos si es porque van a abandonar el programa, porque no les interesa, porque no son conscientes de que disponen de una serie de ventajas acumuladas, pero sí sabemos que desde el punto de vista de los patrocinadores del programa suponen clientes que pueden no ser fieles, en mayor o menor medida tienen riesgo de abandonar ese programa de fidelización. Nos fijaremos sobre todo en el patrón de redención de puntos. Veremos que se trata de uno de los indicadores básicos de la eficacia del programa: si los clientes pueden conseguir puntos es porque luego van a poder intercambiarlos por regalos que les resultan interesantes. Sin la redención de puntos no se entiende la existencia de un programa de fidelización.

No sabremos estos motivos, pero sí podremos saber algunas características básicas que nos permitan clasificar a los clientes en aquellos que nos son de mayor interés de los que no nos interesan. Y una vez establecidos los objetivos de nuestra campaña de "recuperación", se tratará de contactar con esos clientes para saber cuáles son sus motivos, sus preferencias con respecto al programa de puntos, por qué no los intercambian, etc. Esta labor de "detección" de los clientes a los que se va a proceder a un complejo y normalmente muy costoso programa de relación requiere, al igual que la detección en el campo médico, de unos procedimientos suficientemente estables y potentes como para permitir una evaluación rigurosa, y un esquema completo de análisis coste-beneficio. Defenderemos en esta tesis que ese enfoque es el de las curvas ROC.

El campo del CRM tiene su propio argot, y creemos muy importante establecer sus objetivos, sus motivaciones, su posición actual en la actividad empresarial, antes de abordar los aspectos más metodológicos. Por esta razón, el capítulo 2 presentará el estado actual de la cuestión de la fidelidad en CRM, y la razón de ser y principios fundamentales de las tarjetas de fidelización. Todo ello realizado a partir de una reciente revisión de la literatura de Marketing, principalmente anglosajona, como suele ser habitual.

El tercer capítulo traza la historia y los campos de aplicación de las curvas ROC, o "análisis ROC", sin entrar en una explicación detallada de los procedimientos, que queda reservado para el capítulo 4.

Las curvas ROC se han caracterizado siempre por ser un área de investigación llevada por las aplicaciones. Una de las figuras clave en su desarrollo y expansión es J. A. Swets. Leyendo cualquiera de sus libros o artículos nos daremos cuenta de que intenta plantear ejemplos reales, evita sistemáticamente los típicos ejemplos o datos simulados típicos del campo de la estadística. Algunos de sus artículos son muy difíciles de seguir, por la cierta complejidad de algunas de las bases teóricas de las curvas ROC, pero no falta en ellos un ejemplo real, o varios. Estos ejemplos se nota que han estado trabajados "a mano", no se presentan prolijas salidas de programas estadísticos, o largas tablas de resultados, sino sintéticas curvas ROC, o su representación en "gráficos binormales". Éstos últimos son representaciones en los que las coordenadas son puntuaciones típicas

de la distribución normal, o el área que dejan por debajo. En vez de una curva, lo que producen (supuestas distribuciones normales subyacentes) son rectas. Muchos de los gráficos que se presentan en los artículos especifican que se han hecho en "papel de distribución normal", que parece ser estaba disponible para el trabajo de los estadísticos hace tiempo. Aunque parezca una curiosidad, dice mucho del carácter "artesanal" y sintético que marca sin duda el desarrollo de las curvas ROC.

Otro de los pilares fundamentales de la expansión de las curvas ROC en el campo del diagnóstico médico ha sido Charles Metz, quien propuso ya en los 80 métodos prácticos para estimar las curvas ROC, según el "modelo binormal", con programas informáticos que permitían su cálculo independientemente de los procedimientos manuales de dibujo anteriores. Seguíamos teniendo los mismos supuestos difíciles de cumplir, pero se podía sistematizar la estimación de curvas en diferentes ámbitos, típicos del campo de trabajo de Metz, el diagnóstico por la imagen.

En el año 1982 se publica un artículo básico, el de Hanley y Mc Neil, en *Radiology*. Se trata de una revista que como su nombre indica leen principalmente los radiólogos, y por cierto, difícil de encontrar. Este artículo explica mejor que ninguno hasta el momento la potencia de las curvas ROC en el campo de la evaluación de sistemas diagnósticos, y sobre todo plantea un método no paramétrico (no requiere supuestos sobre la forma de la distribución) para la comparación de áreas bajo la curva, la medida global de precisión diagnóstica. Este procedimiento, en todo similar al contraste W de Wilcoxon, estaba preparado para datos no relacionados (grupos independientes), y para estudios de evaluación de sistemas diagnósticos mediante escalas de valoración. En 1983 presentan el mismo enfoque para datos relacionados (un mismo conjunto de "lectores" que evalúan varios instrumentos diagnósticos). Desde ese momento comienza una enorme difusión de estas técnicas en la Medicina clínica, no sólo, pero principalmente, en la Radiología y Diagnóstico por la imagen. Recordemos que en esos tiempos no había ordenadores personales, y predominaban los cálculos a mano o mediante calculadoras. El estudio de estos artículos ha retrotraído a quien esto escribe a los tiempos en que los libros publicaban estos procedimientos, pero que exigían una comprensión muy buena de la técnica que se estaba manejando, porque cualquier pequeño error en el proceso producía enormes cambios en los resultados. El análisis ROC, con su simple pero potente punto de partida de la tabla de clasificación de doble

entrada, ofrecía esa facilidad de comprensión necesaria para su uso por mucha gente para quienes la estadística era una herramienta para su trabajo.

En 1988 aparece otro artículo clásico, el de DeLong, DeLong y Clarke-Pearson en *Biometrics*, quienes proponen una metodología general para la comparación no paramétrica de las áreas bajo las curvas ROC, basado en el procedimiento del contraste U de Mann-Whitney. También hacen pública una macro para SAS (uno de los paquetes clásicos de estadística, quizá el líder en el campo médico) que por fin permite hacer contrastes estadísticos de igual modo que se hace una *t* de *Student*. Este artículo aparece citado en cientos de otros artículos en el campo médico.

A pesar de esto, y de una creciente aplicación en otros muchos campos además del clínico, como veremos en el capítulo 3, no es hasta prácticamente el año 2003 que un módulo completo de análisis de curvas ROC aparece en un programa de análisis estadístico. Este programa es el NCSS 2004. Su autor principal, Hintze, dice en el excelente manual que lo acompaña, que una de las dificultades más importantes que ha encontrado para el desarrollo de este módulo es la carencia de fuentes de calidad para saber cómo implementar los diferentes enfoques y procedimientos. Efectivamente, hasta el libro de Zhou *et al.* (2002), no ha habido una fuente única para estudiar los procedimientos. Los análisis ROC se usaban en todas partes, pero cada uno citaba fuentes dispersas, o "se lo montaba" un poco a su modo.

El capítulo 4 muestra en detalle todos los aspectos metodológicos de las curvas ROC, en particular de sus dos modelos o enfoques, el paramétrico y el no paramétrico. Mostraremos que las curvas ROC, frente a otras técnicas con objetivos similares, presentan las siguientes ventajas (Swets, 1982, 1986, 1988):

1. Proporcionan un índice de eficacia diagnóstica puro, cualquiera que sea el criterio o punto de corte en el indicador cuantitativo en que se basa la decisión, e incluso independientemente de que dicho indicador esté sesgado. Esta medida resulta extremadamente útil para poder elegir una técnica entre varias en competencia.
2. Estima la probabilidad de diferentes resultados en la tabla de toma de decisiones. De este modo, permiten aislar causas de error en sistemas diagnósticos (por ejemplo, falsas alarmas y omisiones).

3. Proporciona un indicador de criterio de decisión (punto de corte o umbral para la toma de decisión), que permite incluir probabilidades (incluso estimaciones subjetivas de la probabilidad) y costes o utilidades. Este método permite establecer razonadamente reglas o mecanismos óptimos de toma de decisión dentro de un sistema de diagnóstico.

Hemos de decir que las curvas ROC empíricas, bajo ese nombre u otro, se usan desde hace mucho para la evaluación de modelos estadísticos, como la regresión logística (módulo BMDP-LR), o lo que el paquete estadístico SPSS incorporó como "curva COR", en un imposible intento por llevar al castellano el concepto. El objetivo de esta tesis va mucho más allá de estos enfoques descriptivos, utilizando toda la potencia de las curvas ROC binormales cuando sea posible, pero también haciendo uso de todas las capacidades de análisis que vienen incorporadas por defecto en el programa NCSS 2004.

El capítulo 5 de esta tesis es un poco más académico, pero creemos que cumple un papel importante al mostrar cómo el enfoque ROC unifica, en el sentido de que incluye y a la vez mejora, una gran variedad de procedimientos de evaluación de sistemas diagnósticos. Se basa en un artículo también clásico de J.A. Swets de 1986, pero añade otros índices provenientes de la informática, sobre todo de la recuperación de información, y que mostramos (que no demostramos) su similaridad, o se podría decir en algún caso identidad, con los propuestos por las curvas ROC.

Precisamente una de las principales motivaciones de esta tesis es mostrar la utilidad de estas técnicas para los recientes enfoques de "minería de datos". Como explicaremos brevemente en el capítulo 2, y de lo cual mostraremos un ejemplo en el capítulo 3, la aparición de potentes algoritmos informáticos ha fomentado una enorme aplicación en CRM. Una de las grandes ventajas de estos algoritmos es que no tienen supuestos que cumplir, y así dejamos atrás las típicas dudas de los analistas (¿son mis distribuciones normales? ¿Homoscedásticas? ¿Cómo podemos corregirlos, transformarlos, quitarles casos extremos? ¿Qué hacemos si tenemos muchos casos "*missing*"?). También tenían grandes ventajas en presentar las reglas en lenguaje natural (los árboles de decisión, porque los resultados de las redes neuronales son francamente difíciles de entender).

Pero ¿hasta qué punto son mejores que los tradicionales modelos estadísticos? Y aquí, cualquiera que haya leído artículos sobre este tema, se encontrará la enorme variedad, o sencillamente la nada más allá de el gráfico de ganancias. ¿Prevalencia? ¿Capacidad predictiva? ¿Contraste estadístico entre los resultados de clasificación de varios procedimientos? Nada de esto es conocido en el campo de la minería de datos, con muy contadas excepciones.

El capítulo 5 sirve entonces para justificar nuestra propuesta de utilizar el análisis de curvas ROC como el mejor enfoque para evaluar la "precisión diagnóstica", ¿o deberíamos decir "capacidad predictiva" de los procedimientos estadísticos?

El capítulo 6 entra ya de lleno en el análisis de datos de un "*data mart*" de clientes inactivos de la tarjeta multipatrocinador Travel Club. Nuestro objetivo será identificar los clientes con mayor riesgo de convertirse en "inactivos definitivos" (dejar de utilizar definitivamente la tarjeta y por tanto abandonar en la práctica el programa), "detectar" en otras palabras más cercanas al uso de curvas ROC, los clientes que serán más beneficiosos según los principios que nos marquemos, que deberá incluir un análisis coste-beneficio. Éste es un escollo importantísimo, puesto que no tenemos datos reales, por lo que nos tendremos que conformar con realizar simulaciones de varios escenarios. De cualquier modo, aun incluso conociendo estos datos, será imprescindible utilizar esta técnica de simulación.

Los modelos estadísticos que utilizaremos para la predicción y comprensión de este fenómeno son los tradicionales de regresión logística (Hosmer y Lemeshow, 1989), hasta modelos de minería de datos (*Data Mining*), en particular, los conocidos como "árboles de decisión". Para éstos últimos veremos la utilidad de las técnicas de curvas ROC para evaluar la eficacia predictiva de los diferentes modelos.

Probaremos la eficacia de estas ventajas en los modelos empírico y binormal de la curva ROC. En función de los resultados evaluaremos su propuesta como metodología de amplia aplicación en Marketing en el capítulo de discusión y conclusiones.