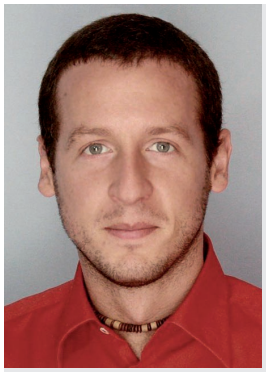
M.6. Implicaciones éticas de la minería de datos

Por **Jorge Franganillo**

3 agosto 2009

Franganillo, Jorge. "Implicaciones éticas de la minería de datos".
Anuario ThinkEPI, 2010, v. 4, pp. 320-324



Resumen: *La minería de datos es el proceso de analizar gran cantidad de datos para descubrir patrones de comportamiento y predecir acciones futuras. La explotación de datos tiene muchas aplicaciones útiles, pero también tiene un enfoque meramente exploratorio que hace discutible la validez de ciertas deducciones. El uso de información personal con fines predictivos tiene consecuencias directas sobre la vida de las personas y exige por tanto actuar en un marco de responsabilidad. Se hace necesario entonces un código de ética.*

Palabras clave: *Minería de datos, Ética, Información personal.*

Title: *Ethical implications of data mining*

Abstract: *Data mining is the process of analyzing large data sets to discover behaviour patterns and predict future actions. Data mining has many useful applications, but also has an exploratory approach that makes questionable the validity of certain deductions. Using personal information for predictive purposes has a direct impact on the lives of people and therefore requires acting in a framework of liability. A code of ethics is necessary.*

Keywords: *Data mining, Ethics, Personal information.*

Discriminación

CIERTOS EXPERTOS pueden describir la conducta de un conjunto de personas basándose en los registros digitales de lo que hacen. La descripción es detallada: qué hacen, qué compran, cómo trabajan, con quién se relacionan.

Es la minería de datos, que suele usarse para discriminar en positivo: al saber por ejemplo qué hábitos de compra tiene un determinado colectivo, es posible orientarle más efectivamente una campaña publicitaria.

Pero también puede usarse para discriminar en negativo: el análisis del registro del correo electrónico de los empleados de una empresa permite identificar a quienes están alimentando redes informales y, en consecuencia, los directivos podrían cambiar la actitud hacia aquéllos.

Un estudio observa que quienes compran coches rojos en Francia son más propensos a incumplir el pago de los créditos (**Chakrabarti**, 2008); esto podría modificar las condiciones crediticias de quienes escogen el rojo para el coche. Suele clasificarse a las personas según estereotipos que se basan en correlaciones estadísticas, pero éstas implican los errores de toda generalización, y así pagan unos por otros.

Antes de ceder información personal, todos deberían saber para qué la van a usar, pero esta cesión puede ser una condición ineludible para que un trámite prosiga su curso. Es evidente por tanto que la minería de datos necesita un código de ética.

Propiedad

Todo individuo es titular de los datos que le conciernen y le afectan personalmente. Así lo establecen la *Directiva 95/46/CE del Parlamento Europeo y del Consejo* y la *Ley orgánica 15/1999 de protección de datos de carácter personal*. Pero estos datos no están en manos de la persona que los genera, sino que están en ficheros ajenos. El individuo es dueño de algo que no controla y entonces ya no es tan dueño. Los dueños de estos datos pasan a ser quienes tienen capacidad tecnológica para recopilarlos y explotarlos.

Las empresas que aprovechan la información procedente de la minería de datos presuponen que el individuo les cede la información que generan con la tarjeta de crédito, el consumo telefónico, etc. Lo presuponen, puesto que dejan constancia escrita –aunque en letra pequeña– de que pueden hacer uso de esa información si el usuario no indica lo contrario. El usuario puede indicar lo contrario, efectivamente, pero para

Figura 1. http://www.ucam.edu/informatica/

ello en general debe escribir una carta y enviarla por correo a la sede de la empresa, y esto implica cultura, tiempo y gasto. El código de ética –queda visto una vez más– es necesario.

"La recopilación de información personal es preocupante porque se realiza de forma deliberadamente silenciosa"

Uso y abuso

La vida cotidiana está digitalizada: cada clic en internet, cada llamada o cada mensaje de móvil, cada compra pagada con tarjeta, todo queda registrado en un sinfín de ficheros. Es un rastro digital que permite dibujar el perfil de las personas y saber qué compran, qué les gusta, con quién hablan, dónde viajan y dónde están, o qué consultan en la Red (**Grau**, 2009). Así, la actividad de miles y miles de personas, incluso millones, produce una masa inmensa de información valiosa que, mediante el adecuado procesamiento matemático, permite identificar costumbres y preferencias.

Al agrupar las personas según los rasgos que comparten y conocer los hábitos del conjunto es fácil suponer qué harán sus integrantes y deducir cómo se puede influir mediante políticas publicitarias, empresariales o de otro tipo que se dirijan específicamente a cada conjunto.

Tanta es esta información, tan valiosa y tan dispersa que ya hay especialistas en rastrear a gran escala las pistas que cada persona, quizá sin saberlo, deja registradas en una serie de gigantescos ficheros. Estos especialistas son matemáticos, programadores, expertos en explotación de datos que, con la ayuda de psicólogos, lingüistas y sociólogos, excavan montañas de datos para extraer el mineral de la información. La minería de datos es el proceso de analizar gran cantidad de datos para descubrir patrones de comportamiento.

Es una actividad de vasto alcance y ayuda a aumentar beneficios (en empresas, comercios, bancos, aseguradoras, etc.), mejorar el diagnóstico y la prevención de enfermedades (en epidemiología, genética, etc.) o velar por la seguridad (contra el terrorismo y el fraude). Los fines son lícitos, pero debe tenerse presente que el tratamiento de la información tiene implicaciones éticas cuando se trata de datos sobre personas.

"La campaña electoral de Barack Obama compró información sobre las inquietudes y los miedos de ciudadanos indecisos"

La recopilación de información personal es preocupante porque se realiza de forma deliberadamente silenciosa. Y el ser humano suele ignorar los peligros que no le son evidentes: cree tener su vida bajo control sin tener presente que personas ajenas a su entorno toman ciertas decisiones que les afectan, basándose en datos personales que no ha proporcionado de manera consciente, o que creía olvidados o secretos (**Garriga**, 2004).

Objetos cotidianos como la tarjeta de crédito o la del supermercado generan una información valiosa que permite afinar las estrategias de marketing y hacerlas más efectivas que las tradicionales, porque ya se sabe qué compra y cuánto gasta un determinado conjunto de clientes. Los supermercados analizan las compras para detectar asociaciones entre los artículos. Los que se suelen comprar juntos, ¿deben estar próximos entre sí, para priorizar la comodidad del cliente, o separados para prolongar su permanencia en el establecimiento e incitarlo así a que haga compras imprevistas?

La minería de datos revela cómo se puede influir sobre las personas y cómo se las puede manipular para obtener un beneficio que no suele ser mutuo, sino exclusivo de quien posee y explota esos datos.

Las empresas de telefonía móvil registran información valiosa sobre cada abonado: dónde está, a dónde viaja, a quién llama, cuánto gasta. Pueden saber si un cliente es líder de un grupo social o si tiende más bien a quedarse al margen. Hasta pueden detectar clientes descontentos, y hacerles llegar ofertas tentadoras para que renueven el contrato. La minería de datos debería

Figura 2. http://historyofeconomics.files.wordpress.com/

servir para mejorar el servicio de todos los abonados y no sólo para retener al descontento.

La campaña electoral de Barack Obama compró información sobre las inquietudes y los miedos de un numeroso grupo de ciudadanos cuyo denominador común era la indecisión. Examinar esta información les permitió agrupar la población en varias "tribus de valores" para hacer un ejercicio histórico de *micro-targeting* político: los mensajes electoralistas fueron más efectivos porque iban dirigidos específicamente a los votantes indecisos (**Baker**, 2009a). Sin ánimo de poner en duda la honestidad de la política electoral estadounidense, es evidente que el objetivo de esta táctica tiene algo de manipulador.

La minería de datos permite hacer más productivos a los empleados a fuerza de controlar su actividad. En un ambicioso proyecto, la multinacional *IBM* ha analizado el comportamiento de cincuenta mil técnicos para extraer modelos matemáticos: se trataba de inventariar las habilidades de cada uno para averiguar matemáticamente el modo más rentable de utilizarlas (**Baker**, 2009b).

También permite detectar qué empleados buscan un ascenso o cambiar de empresa. Los registros del navegador de internet y del correo electrónico contienen los patrones de comunicación de cada trabajador, una información a la cual el jefe puede acceder. ¿Es aceptable que una empresa revise las comunicaciones de sus empleados?

En 2007, el *Tribunal Supremo* de España dictó sentencia: los servicios que se ponen a disposición de los trabajadores merecen protección, pero los empresarios pueden rastrear mensajes e historiales si antes avisan de que las comunicaciones serán supervisadas (**Sahuquillo**, 2007).

Los expertos en explotación de datos pueden sumergirse en un océano de información en busca de patrones que definan la conducta terrorista. Pero esa conducta es un misterio e incita a formular suposiciones y construir hipótesis, quizá inciertas, sobre qué constituye un movimiento sospechoso (**Baker**, 2009c). La minería de datos no es entonces adecuada para descubrir comportamientos terroristas. Utilizarla en la seguridad nacional y en la aplicación de la ley sería con-

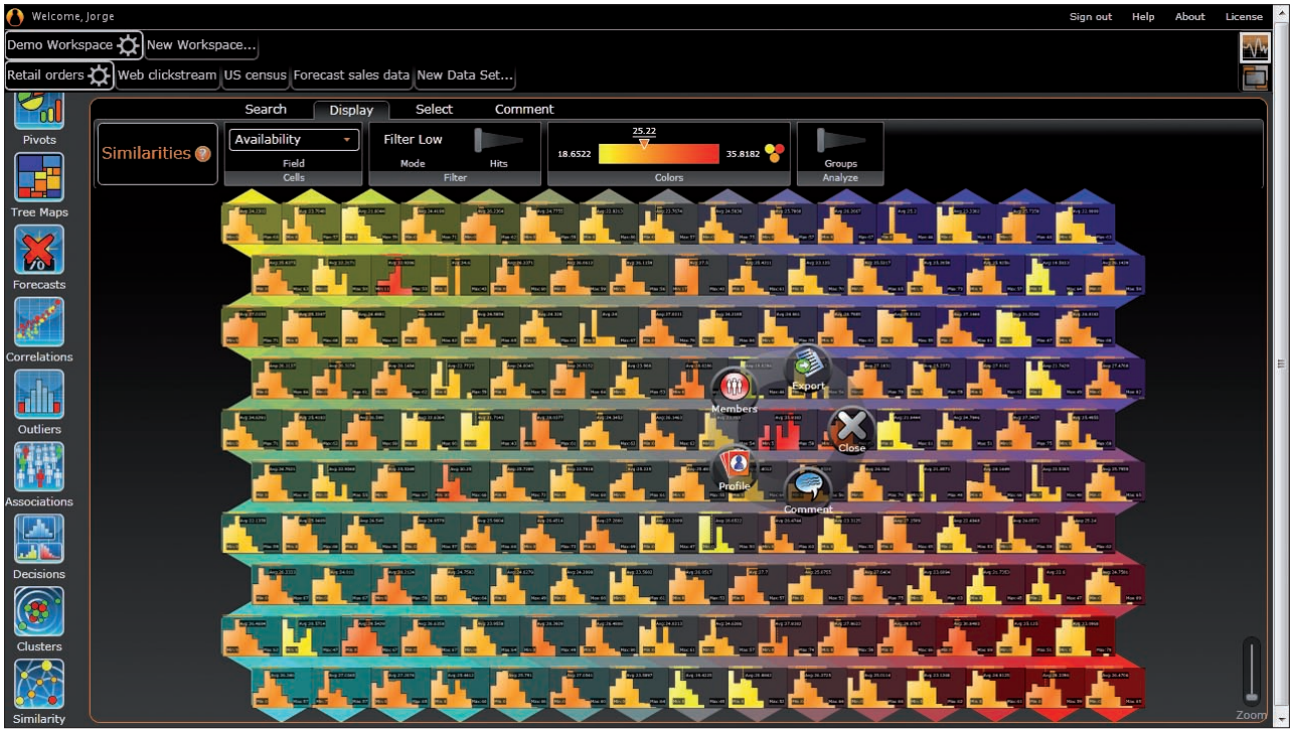


Figura 3. Data Applied software 1

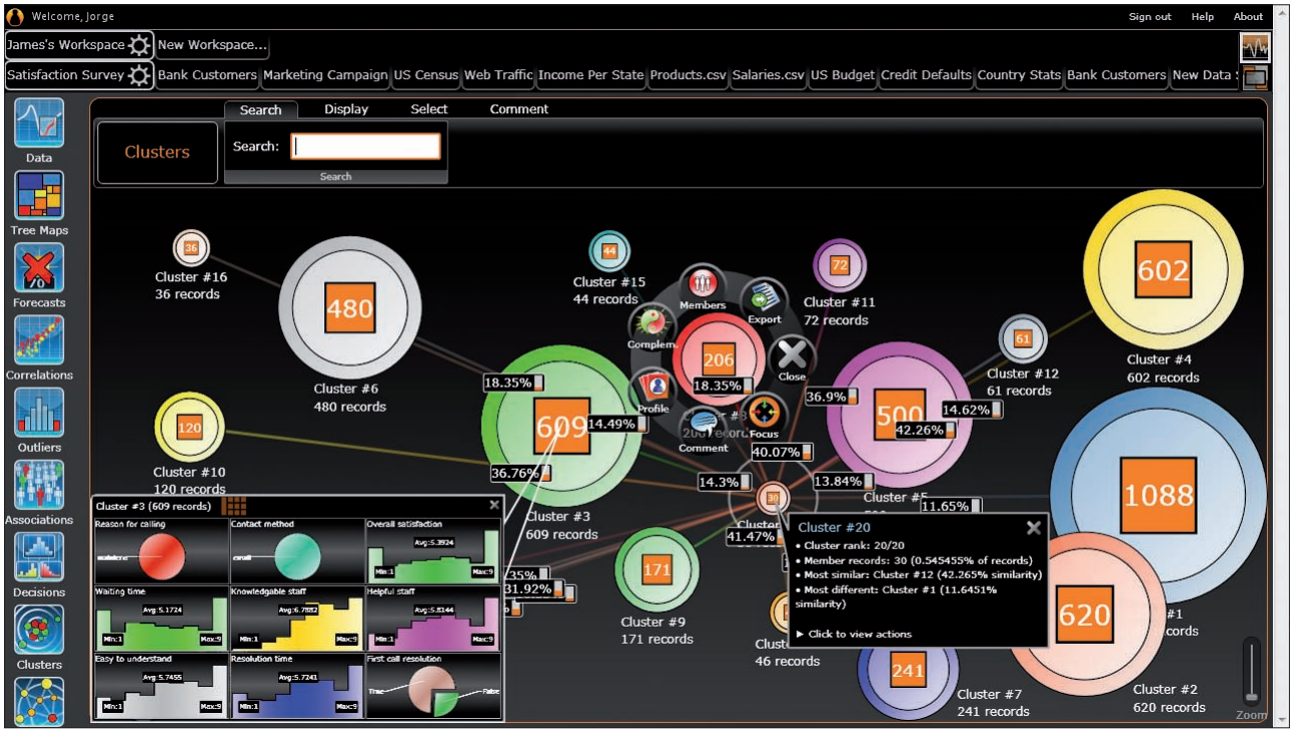


Figura 4. Data Applied software 2

traproducente: supondría malgastar el dinero de los contribuyentes para vulnerar la privacidad e infringir libertades civiles. Y la seguridad jamás debe convertirse en un pretexto para imponer vigilancia y recortar libertades. Se necesitará entonces una profunda regulación para proteger los derechos y la identidad de las personas, y evitar entrar en una sociedad de la vigilancia que, a pesar del control, sea incapaz de mantenernos seguros.

En internet, la minería de datos ayuda a mejorar la usabilidad de las sedes web mediante el análisis del proceder de los visitantes. Y también ha servido para mostrar que una web tan popular como *Facebook* es un sitio inseguro (**Kelly**, 2008). Esta red social ya acumula numerosas críticas sobre cómo maneja los datos de los usuarios, que se ven desprotegidos. La empresa de **Mark Zuckerberg** ha sido legalmente cuestionada porque retiene los datos de los usuarios que han solicitado darse de baja y no los protege como debe cuando los cede a terceros (**Shields**, 2008; **Noain**, 2009).

Mientras no se resuelven estos problemas, los usuarios deben ser cautos al introducir sus datos en redes sociales: cuantas más informaciones faciliten, más expuestos estarán a usos que puedan atentar contra su privacidad.

Acción y reacción

Las personas no permanecen impasibles ante esta incómoda realidad. Y precisamente en internet, los buscadores afrontan esta situación constantemente: hay toda una industria dedicada a modificar las páginas web para mejorarles la posición en la lista de resultados, porque una página pierde visibilidad si no tiene buena puntuación, y en un mundo saturado de mensajes la visibilidad es importante. Entonces, las personas usan los mecanismos de la minería de datos para manipular empresas cuyo funcionamiento está guiado por la misma minería.

Esta minería es útil y necesaria, y las aplicaciones en astronomía y meteorología son ejemplos de cuánto nos pueden facilitar la vida cotidiana. Pero surge rápidamente el fantasma de la discriminación cuando se aplica a las personas y aporta argumentos –o sólo sospechas– basadas en datos estadísticos, para denegar un crédito, rechazar una solicitud, o para identificar quién recibirá una oferta especial y quién deberá pagar el precio estándar.

"La significación de unos datos no proviene de la capacidad técnica para organizarlos, sino de la intencionalidad y, por tanto, de los prejuicios con que el profesional los gestiona"

Aunque no está en un vacío legal, la minería de datos necesita un código de ética porque la evidencia del día a día indica que no toda

recopilación de datos se lleva a cabo con medios aceptables desde el punto de vista ético. Ni todo el uso de la información que se obtiene persigue sanos objetivos.

Considerando que la discriminación puede ser negativa, que puede carecer de ética e incluso que puede ser ilegal, ¿en qué circunstancias es lícito discriminar? Todo apunta a que la legitimidad de la minería de datos depende de cómo se aplique.

En medicina es lícito tener información sobre características raciales o sexuales si esta información queda restringida al uso médico, pero sería ilícito usar tales variables para analizar el comportamiento en la devolución de préstamos, por ejemplo. Pero incluso cuando se excluye esta información sensible, hay cierto riesgo de que los modelos se construyan basándose en variables que equivalen a condición racial, religiosa o sexual. Un dato aparentemente inexpresivo como el código postal puede ser un factor de discriminación negativa si va asociado a una identidad étnica, como ocurre en algunos distritos de muchas ciudades.

Es difícil quedar fuera del radar. La vida moderna sólo es posible si se usan los instrumentos que para ella se proponen, y éstos permiten primero el rastreo y después la explotación de grandes conjuntos de datos personales. Estos instrumentos nos hacen la vida más fácil, más cómoda y más segura, pero tienen un precio: nos exponen a prácticas de dudosa ética que pueden hacernos perder parcelas de libertad y privacidad.

Precisamente para no perder libertad y privacidad, es necesario saber cómo se usará la información que generamos cotidianamente, cómo se protegerá su confidencialidad e integridad, qué consecuencias pueden derivarse y qué derecho tenemos a rectificar o incluso retirar la información que nos atañe. Todo debe explicarse en un lenguaje comprensible –y no en una jerga legal de letra pequeña– y desde un principio, puesto que la minería puede ir luego más allá de la finalidad para la cual se recopilaron los datos. La tenencia de éstos no confiere el derecho a usarlos con objetivos distintos de los previstos de manera explícita.

Se puede cuestionar que circule tanta información personal. Ésta es un patrimonio que reclama una protección especial puesto que la simple acumulación de informaciones puede volverse un acto especulativo, sobre todo cuando se poseen dispositivos técnicos para darles significado e interpretarlos, quizá incorrectamente.

La significación de unos datos no proviene de la capacidad técnica para organizarlos, sino de la intencionalidad y, por tanto, de los prejuicios con que el profesional los gestiona. Cuando la información resultante es susceptible de un uso ilegítimo, surge la necesidad de marcar los límites de la actuación de las personas o instituciones responsables de manejar esos datos.

Referencias

Baker, Stephen. "What data crunchers did for Obama". *Business week*, 2009a.
http://businessweek.com/technology/content/jan2009/tc20090123_026100.htm

Baker, Stephen. "Data mining moves to human resources". *Business week*, 2009b.
http://businessweek.com/magazine/content/09_12/b4124046224092.htm

Baker, Stephen. *Numerati: lo saben todo de ti*. Barcelona: Seix Barral, 2009c.

Chakrabarti, Soumen et al. *Data mining*. San Francisco: Morgan Kaufmann, 2008.

Garriga, Ana. *Tratamiento de datos personales y derechos fundamentales*. Madrid: Dykinson, 2004.

Grau, Abel. "Tus datos íntimos son una mina". *El país*, 3 junio 2009.
http://elpais.com/articulo/sociedad/20090603elpepisoc_1/Tes

Kelly, Spencer. "Identity 'at risk' on Facebook". *BBC news*, 1 mayo 2008.
http://news.bbc.co.uk/2/hi/programmes/click_online/7375772.stm

Noain, Idoya. "Ultimátum de Canadá a Facebook para que garantice la intimidad". *El periódico de Catalunya*, 26 julio 2009.
http://elperiodico.com/default.asp?idnoticia_PK=632529

Sahuquillo, María R. "Ordenadores sin secretos para el jefe". *El país*, 4 noviembre 2007.
http://elpais.com/articulo/sociedad/20071104elpepisoc_1/Tes

Shields, Maggie. "Facebook viola la privacidad". *BBC Mundo*, 31 mayo 2008.
http://news.bbc.co.uk/hi/spanish/science/newsid_7428000/7428904.stm