



Facultad de Ciencias Económicas y Empresariales

OPORTUNIDADES ESTRATÉGICAS Y RIESGOS EN LA IMPLEMENTACIÓN DEL *BIG DATA* EN EL SECTOR FINANCIERO ESPAÑOL

Autor: Pablo López de Miguel

Director: Patricia Candelaria Soriano Machado

RESUMEN

El *big data*, o la gestión de macrodatos, se ha convertido en un elemento esencial de todo negocio, sobre todo en el sector financiero y la banca a nivel mundial. Cada vez se utiliza más el *big data* para la gestión de los macrodatos de clientes de bancos, aunque muchas veces aparece desapercibido. Ya en 2015 el informe “*big data* en el sector financiero español”, elaborado conjuntamente por las consultoras Ernst and Young, FrontQuest y Teradata, muestra como alrededor de un 30% de las empresas del sector financiero (banca y aseguradoras) en España, aprovechan ya en mayor o menor medida, las oportunidades que ofrece esta tecnología (EY.com, 2015) y este número solo ha crecido desde entonces.

Esta nueva tecnología ofrece enormes oportunidades estratégicas para estas entidades, como la mejora del servicio al cliente y una gestión de riesgos y fraudes aún más eficiente y eficaz. Sin embargo, estas oportunidades también van acompañadas de riesgos y barreras que las entidades financieras españolas deberán de superar, ajustándose a las restricciones legales y a las limitaciones actuales de la tecnología.

Palabras clave: *big data*, macrodatos, segmentación, *online*¹, rentabilidad, tecnología, riesgos, *start-up*², *fintech*³.

¹ En línea.

² Empresa emergente.

³ Industria financiera que aplica la tecnología para mejorar los servicios financieros.

ABSTRACT

Big data, or the management of large amounts of data, has become an essential element of any business, especially in the financial sector and banking worldwide. Big data is increasingly used for the management giant databases from bank clients, although it often appears unnoticed. Already in 2015, the report “*big data* en el sector financiero español”, developed jointly by the consultants Ernst and Young, FrontQuest and Teradata, shows how around 30% of companies in the financial sector (banking and insurance) in Spain take advantage of, to a greater or lesser extent, the opportunities offered by this technology (EYy.com, 2015) and this number has only grown since then.

This new technology offers enormous strategic opportunities for these entities, such as improving customer service and even more efficient and effective risk and fraud management. However, these opportunities are also accompanied by risks and barriers that Spanish financial institutions must overcome, in accordance with legal restrictions and current limitations of technology.

Key words: big data, targeting, segmentation, online, profitability, technology, risks, start-up, fintech.

ÍNDICE

1. INTRODUCCIÓN.....	1
1.1 OBJETIVOS	1
1.2 METODOLOGÍA.....	1
1.3 ESTADO DEL <i>BIG DATA</i> EN EL ENTORNO EMPRESARIAL.....	2
1.4 PARTES DEL TFG	5
2. OPORTUNIDADES DEL <i>BIG DATA</i> EN EL SECTOR FINANCIERO.....	6
2.1 SERVICIO AL CLIENTE.....	6
2.1.1 <i>Herramienta de fidelización</i>	7
2.1.2 <i>Segmentación de clientes</i>	8
2.1.3 <i>Marketing y targeting avanzado para empresas</i>	11
2.1.3.1 <i>Análisis de la información del weblog</i>	12
2.1.4 <i>Gestión de multicanalidad</i>	12
2.1.5 <i>Asesoramiento financiero</i>	14
2.2 GESTIÓN DE RIESGOS.....	15
2.2.1 <i>Riesgo de crédito</i>	15
2.2.2 <i>Otros riesgos</i>	17
2.3 GESTIÓN DE FRAUDES	18
2.3.1 <i>Detección de fraude en los medios de pago</i>	18
2.3.2 <i>Prevención de blanqueo de capitales y financiación del terrorismo</i>	20
3. PRINCIPALES BARRERAS PARA SU APLICACIÓN.....	21
3.1 PROTECCIÓN DE DATOS	21
3.1.1 <i>Marco jurídico de la Protección de Datos</i>	21
3.1.2 <i>Riesgos para el big data</i>	23
3.1.2.1 <i>Anonimización y el riesgo de reidentificación</i>	23

3.1.2.2 Consentimiento informado	25
3.1.2.3 Riesgo en la toma de decisiones automatizadas	28
3.1.2.4 Problema de la propiedad de datos	29
3.2 CALIDAD DE LOS DATOS	30
3.2.1 <i>Error por azar</i>	31
3.2.2 <i>Error por confusión</i>	34
3.3 CAPACIDAD DE LA TECNOLOGÍA	35
3.4 FALTA DE CONOCIMIENTO ESPECIALIZADO	36
3.5 CONFIANZA DE LOS CLIENTES (DESCONFIANZA EN LAS ENTIDADES)	37
4. CONCLUSIONES.....	38
BIBLIOGRAFÍA	41

ÍNDICE DE GRÁFICOS

GRÁFICO I: VARIABLES 1 A 5	32
GRÁFICO II: 1000 VARIABLES	32
GRÁFICO III: VARIABLES 30 Y 132	33
GRÁFICO IV: NIVEL DE CONFIANZA DE LA POBLACIÓN ESPAÑOLA EN SUS INSTITUCIONES FINANCIERAS Y EN SUS BANCOS ENTRE 2005 Y 2012	37

1. Introducción

1.1 Objetivos

En España se ve como la demanda, inversión y uso de la tecnología *big data* continúa creciendo, sin embargo, las entidades solo son capaces de sacarle poca rentabilidad real. El presente trabajo de investigación académica está orientado en analizar dos aspectos de singular importancia para el desarrollo del *big data* en el sector financiero de español, (i) cuales son las posibles aplicaciones para esta tecnología y como pueden las entidades que lo implementan obtener rentabilidad, y (ii) cuáles son los riesgos y las barreras que están frenando su desarrollo e implementación y como pueden estas entidades superarlas.

1.2 Metodología

Para conseguir el objetivo de este trabajo de investigación se empleará numerosas fuentes de información. Utilizando estudios y artículos desarrollados por expertos en la materia, al igual que encuestas llevadas a cabo por diversas consultoras e instituciones financieras.

Muchos de los ejemplos utilizados en el proyecto provienen de pequeñas empresas, *start-ups*, y *fintechs* que verdaderamente están impulsando la utilidad y la experimentación con esta tecnología y otros muchos ejemplos, de las grandes instituciones de España (BBVA y el Banco Santander), al ser los dos bancos más grandes de España, que más recursos están empleando en esta materia y que, además, llevan casi cuatro años de ventaja en el estudio e implementación de esta tecnología (Martínez, 2017).

Por último, la información sobre los riesgos técnicos y legales del *big data* se desarrolla a partir de documentos oficiales de la Agencia Española de Protección de Datos, los propios Reglamentos y estudios estadísticos sobre la materia llevados a cabo por especialistas y consultoras.

1.3 Estado del *big data* en el entorno empresarial

Antes de analizar el estado del *big data* en el entorno empresarial, primero se debe conocer que es el *big data* y cuáles son los elementos que lo caracterizan. El *big data* se refiere a la enorme cantidad de datos que los clientes, negocios y organizaciones crean. Estos datos son tan numerosos, complejos y se desarrollan a tal velocidad que es imposible organizarlos y analizarlos con herramientas y software convencionales.

Se obtienen de diversos medios, como el uso de cookies en páginas web, la identificación por radiofrecuencia (RFID), el rastreo por GPS, la gestión de búsquedas en internet y redes sociales, registros de llamadas, el control de movimientos bancarios y el control de datos económicos y de censos ... entre otros muchos datos.

Debido al enorme uso de las nuevas tecnologías (cada vez más comunes en la vida diaria), se están generando enormes cantidades de datos que, si no son organizados y posteriormente analizados, acaban desaprovechados de forma muy poco productiva.

La definición de *big data* se puede dividir en los 5 elementos que lo caracterizan y diferencian de otros conceptos como *Datamining* o *Business Intelligence*⁴ (Serrano Acitores et al., 2019).

Se trata de las 5 Vs del *big data*.

1) Volumen:

La cantidad de datos creados y almacenados no deja de crecer.

Hoy en día ya se tiene que cambiar la dimensión de procesamiento de datos, se ha pasado de terabytes a petabytes (y recientemente a los zettabytes) y en 2020 se prevé que se alcance los 44 zettabytes (Elmer-Dewitt, 2011).

El reto del *big data* está en cómo poder analizar, organizar y aprovechar este ingente número de datos.

2) Variedad:

El *big data* otorga la capacidad de combinar datos de orígenes distintos y formatos no homogéneos, lo cual aumenta la capacidad de análisis de las empresas.

⁴ Ambos son procesos de extracción y análisis de datos, pero a un volumen mucho menor que el *big data* y con información estructurada, a diferencia del *big data*, que es capaz de analizar información estructurada y no estructurada a tiempo real.

3) Velocidad:

Las nuevas tecnologías deben de ser capaces de organizar y analizar los datos de forma rápida, debido a la enorme velocidad a la que se están creando. Cuanto más retraso en el análisis de datos, mayor será el desfase entre los resultados y el mercado actual.

4) Valor:

Los datos por sí solos carecen de significado, la función del *big data* está en analizar y combinarlos con el fin de obtener valor para la entidad.

5) Veracidad

Con esto se refiere a la integridad de los datos. Este es uno de los aspectos más importantes para el desarrollo del *big data*, para que pueda convertirse en un recurso fiable para las entidades financieras españolas.

La gran mayoría de los riesgos y barreras que la tecnología afronta hoy en día, como se verá más adelante en el proyecto, está en reconocer la calidad de los datos y la relación de causalidad entre las variables. Por lo tanto, es necesario un desarrollo mayor de la capacidad de análisis de esta tecnología (en cuanto a la integridad de los datos) para así hacer frente a estas barreras.

Estos datos suponen una enorme ventaja estratégica para aquellas instituciones que sean capaces de analizarlos correctamente.

Al manipular estas enormes cantidades de datos, estas organizaciones pueden elaborar informes estadísticos y modelos predictivos que, no sólo les pueden ayudar en su trabajo, sino también les brindan la oportunidad de ofrecer un mayor rango de mejores servicios para sus clientes.

Este análisis permite a estas instituciones conocer mejor a sus clientes, responder a posibles preguntas que tengan o que ni siquiera sabían que tenían, así como conocer mejor su negocio y el de su competencia. Pueden analizar tendencias que le permitan llevar a cabo su negocio de una manera más eficiente, así como identificar nuevas oportunidades de negocio.

No aprovechar estos datos supone una oportunidad perdida para la empresa y un desaprovechamiento de sus recursos. Es una información disponible para el desarrollo de cualquier empresa, simplemente mediante su recopilación y organización.

Desde el punto de vista del sector financiero son enormes las oportunidades que el *big data* ofrece. Por una parte, desde un punto de vista interno de la institución, el *big data* puede ayudar a localizar patrones de consumo que permitan segmentar a los clientes y llevar a cabo una oferta especializada de productos según sus necesidades. También se puede usar como herramienta de fidelización para los clientes actuales, para gestionar el riesgo de las inversiones de la propia institución y de sus clientes, o para analizar la posición de la competencia, así como para controlar los *KPIs*⁵ (*Key Performance Indicators*) de la propia entidad.

Por otra parte, desde un punto de vista más externo, el *big data* y su gestión tiene numerosos usos que pueden también beneficiar a los clientes de estas instituciones. Por ejemplo, se puede utilizar como herramienta de ayuda al cliente y para la gestión de la multicanalidad y, de esta forma, mejorar la comunicación con los usuarios. También puede ofrecer a los negocios miembros del banco un servicio de *targeting*⁶ y marketing avanzado según los patrones de su *target market*⁷ y de su competencia y, por otro lado, ser usado también por el propio banco para mejorar sus estrategias de marketing. Finalmente, se puede detectar y consecuentemente prevenir el fraude de gastos y tarjetas de créditos al analizar los patrones de consumo, al igual que otras muchas variables.

Estos son algunas de las aplicaciones del *big data*, que son importantes para el desarrollo del sector financiero español y que podrán darles a los bancos diversas oportunidades estratégicas frente a su competencia nacional e internacional.

Actualmente, el desarrollo de esta tecnología en España está un poco estancado. Ha habido mucha inversión en estos instrumentos, pero, de momento, se ha obtenido muy poco retorno de dichos fondos.

Las entidades financieras que están impulsando en mayor medida el *big data* y sus capacidades en España son el BBVA y el Banco Santander. Todas las entidades conocen el *big data* y su importancia, pero aún están lejos de saber cómo aplicarlo para sacarle

⁵ Una medida del nivel de rendimiento de un proceso de la entidad.

⁶ Apartado del marketing online cuyo objetivo es relacionar la publicidad con las necesidades del cliente.

⁷ El destinatario ideal de una campaña de publicidad.

partido en forma de rentabilidad. Además, existen numerosas barreras para la implementación de esta tecnología en las empresas financieras españolas, como la actual capacidad tecnológica, la ingente necesidad de recursos, la falta de conocimiento especializado sobre el tema, o la escasa madurez de muchas de estas empresas a nivel de gestión de datos.

1.4 Partes del TFG

El trabajo se estructura en dos partes fundamentales. En la primera se analizan las posibles aplicaciones del *big data* en el sector financiero español, como se están empezando a implementar estas aplicaciones, que ventaja estratégica dan a las entidades que lo utilizan, y como pueden estas entidades sacar rentabilidad de esta tecnología. En la segunda parte se desarrollarán los riesgos y barreras que están impidiendo el desarrollo y la implementación de estas tecnologías, y que pueden hacer las empresas que utilizan esta tecnología *big data* para hacer frente a ellas.

2. Oportunidades del *big data* en el Sector Financiero

A continuación, se realizará un análisis en más detalles algunas de las **aplicaciones** que han sido resaltadas anteriormente, centrándose por supuesto en su actual implementación y desarrollo en instituciones españolas, pero también mirando al panorama internacional para analizar como empresas internacionales están poniendo en práctica la gestión de macrodatos, y cómo puede usar el sector financiero español estos conocimientos para su negocio.

Este análisis se apoyará en el estudio realizado por Julián García Barbosa y publicado en su artículo “El papel del *big data* en tres áreas críticas del sector bancario”, para analizar las posibles aplicaciones (Barbosa, 2013).

En este artículo el autor reconoce tres posibles aplicaciones para el *big data* en el sector bancario, aunque también puede ser aplicado a otras áreas de *financial services*⁸. Estas tres aplicaciones son el servicio al cliente, la gestión de riesgos, y la gestión de fraudes.

2.1 Servicio al cliente

La principal función para la implementación de la tecnología *big data* en las entidades financieras debe ser la de proporcionar un mejor servicio al consumidor, es decir, un **servicio personalizado** y adaptado a las necesidades de cada cliente. Una mayor satisfacción del cliente, junto con una gama amplia de servicios más adecuada a sus necesidades, llevará a una mayor rentabilidad para estas entidades.

Las empresas en el área de *financial services* tienen un enorme acceso a muchísima información del cliente a través de su información personal, sus cuentas, productos y servicios, el uso que el cliente le dan a estos, sus operaciones en cajeros, banca y operaciones on-line y móvil, compras con tarjeta, problemas con servicio de atención al cliente, ... conocer mejor al consumidor supone una enorme ventaja para estas compañías.

Cabe destacar aquí la V de “variedad” entre las Vs de *big data*, ya que es importante que

⁸ El sector de servicios financieros.

la tecnología *big data* sea capaz de combinar correctamente toda la ingente información del cliente para así llevar a cabo un mejor análisis de su situación.

Toda esta gran cantidad de datos del cliente no hacen más que abrir la posibilidad de múltiples aplicaciones del *big data*. A continuación, se llevará a cabo un análisis de las principales, resaltando también como las empresas españolas lo están utilizando.

2.1.1 Herramienta de fidelización

El concepto de fidelización no es algo nuevo para las empresas del sector financiero, aunque sí ha cambiado la forma de abordarlo. Hoy en día, las entidades financieras deben utilizar toda la información de sus clientes que esté a su disposición para desarrollar productos especializados dependiendo de sus propias necesidades. Además, estos datos permiten a las entidades reconocer necesidades desconocidas por los propios clientes para así desarrollar en el futuro nuevos servicios y productos.

Como se ha explicado anteriormente, hoy en día pocas entidades financieras españolas están impulsando y aplicando el *big data* debido a las numerosas barreras a la hora de su implementación. La avanzadilla en el sector en España se encuentra en BBVA y Banco Santander.

En el caso de BBVA, en la actualidad está analizando los datos de sus clientes para crear nuevos productos, utilizando de esta manera el *big data* como herramienta de fidelización. La asesora de *corporate finance*⁹ Abra Invest ya publicó en 2016 un artículo sobre la *startup*¹⁰ Datio (Gutiérrez, 2016).

Esta empresa, formada a partir de la unión de BBVA y Stration, está especializada en el uso de *big data* y permite a sus clientes analizar imágenes, videos, correos electrónicos, comentarios, suscripciones, información relevante, acontecimientos financieros, etc. Datio utiliza esta tecnología para analizar y gestiona los datos del sector bancario proporcionados por BBVA, así como para adaptar los productos de la entidad a las necesidades de los clientes reales (Gutiérrez, 2016).

⁹ Finanzas corporativas.

¹⁰ Empresa emergente.

Como bien explica la empresa en su página web¹¹: “Hoy en día el sector financiero es el que genera un mayor volumen de datos: pagos con tarjeta, operaciones realizadas en los cajeros, productos contratados, pagos con dispositivos móviles, etc. Todos estos datos ofrecen un mayor conocimiento del cliente, y esto permite mejorar los procesos del banco y desarrollar productos y servicios que realmente interesan a cada uno de sus consumidores.” (Datio, 2019).

La fidelización del consumidor se basa en un simple hecho: su **satisfacción**. Si las entidades financieras, a través de la tecnología *big data*, son capaces de conocer y entender mejor las necesidades de sus clientes, serán capaces de ofrecer y desarrollar productos que se adapten a estas necesidades, asegurando así la permanencia del cliente en la entidad.

Un informe elaborado por la multinacional CDC software revela que los bancos y cajas españoles pierden un 12 % de sus clientes anualmente debido, en gran medida, a la insatisfacción del cliente (Rojas, 2011). Esto demuestra la enorme falta de conocimiento que estas empresas tienen de sus clientes. Es más, el estudio revela que de media el 22 % de los clientes no están satisfechos con el servicio o los productos que la entidad ofrece. El 55 % de estos clientes descontentos abandonarán la entidad y contarán su mala experiencia a una media de 10 personas. Por lo tanto, la entidad no solo pierde clientes actuales, sino que además pierde a posibles futuros clientes.

Consecuentemente, la clave para la fidelización del cliente está en su satisfacción y, para obtener esto, las empresas financieras españolas deben conocer mejor a sus clientes y sus necesidades, siendo la tecnología *big data* la que les permita llevarlo a cabo.

2.1.2 Segmentación de clientes

Además de ofrecer mejores productos, los datos sobre clientes también otorgan a estas entidades la capacidad de agrupar a los clientes en diferentes segmentos, cada uno caracterizado por ciertas condiciones que describen el **comportamiento** de los individuos que pertenecen a ellos. Este es un elemento clave que lleva a la fidelización del cliente. Es imposible conocer las características individuales de cada uno de los clientes, pero si

¹¹ <https://www.datio.com>

la organización es capaz de agruparlos según ciertos indicadores de interés para la entidad, será capaz de ofrecer ciertos servicios y productos a estos segmentos.

La segmentación de clientes no es algo nuevo, sobre todo en la industria tecnológica, pero el *big data* está revolucionando el nivel al que se puede caracterizar un cliente. Muchas empresas tecnológicas llevan años utilizando esta tecnología para llevar a cabo una mejor oferta de productos a sus clientes. Ya a finales de los 90, Amazon puso en marcha un equipo para **automatizar** el procedimiento de recomendación de sus productos (Linden et al., 2003). Este sistema utiliza el *big data* para recopilar información sobre el cliente, basado en sus datos personales, compras previas, búsquedas, y muchas más variables. Para segmentar al cliente, hay que hacer un ofrecimiento basado en sus **preferencias** para predecir su próxima compra. Hoy en día una tercera parte de las ventas de Amazon son el resultado de este sistema de recomendación (Linden et al., 2003). Este mismo sistema es el que entidades financieras ya están empezando a implementar en España, ya que no sólo aumenta la rentabilidad de la entidad al ofrecer más y mejores productos para el cliente, sino que también aumenta la satisfacción, ya que son capaces de encontrar con facilidad productos que necesitan y les interesan. Y como se ha visto, la satisfacción del cliente es un elemento esencial para mantener a los clientes actuales, así como para captar a clientes nuevos.

Existen numerosas características por las que segmentar a los clientes, sin embargo, la entidad debe elegir aquellas que mejor se adaptan a su negocio (en este caso en el sector financiero). Las características por las que un segmento compraría o necesitaría un determinado producto, o por las cuales se podría desarrollar una campaña de marketing para ofrecerles ese producto que pueda llamarles la atención según las propiedades del comportamiento de ese segmento, como por ejemplo datos demográficos, geográficos, psicológicos, o el comportamiento de compra.

El sector financiero lleva años realizando una segmentación de sus clientes, pero por lo general es muy poco eficiente. La aplicación de *big data* permite hacer una segmentación más sofisticada a partir de la información de varias fuentes de datos, que puede llegar incluso a la individualización del cliente y a la segmentación **en tiempo real**. Por ejemplo, se puede utilizar la geolocalización asociada a los canales móviles y a su historial de consumo para ofrecer a los clientes descuentos y beneficios en tiempo real en comercios cercanos.

La disponibilidad de estas enormes cantidades de datos permite esta segmentación dinámica que se ajusta constantemente. Kallerhoff (2012) utiliza el análisis de *cluster*¹² sobre el mix de productos de cada cliente bancario, con el fin de posicionar a los clientes en diferentes segmentos (*clusters*), analizar la configuración estática de estos segmentos, y la probabilidad de que un cliente pase de un segmento a otro. Esta predicción se puede utilizar para recomendar productos ajustados al segmento al que el cliente tiene una alta probabilidad de moverse.

Un ejemplo muy claro es el que da el Blog especialista en Ciencia de Datos PredictLand, utilizado también en numerosas empresas: el análisis RFM de clientes: *recency, frequency y monetary*¹³ (Predictland.com, 2018). Mediante este análisis se calcula el tiempo desde la última compra del cliente, la frecuencia con la que se contrata determinado producto y el valor monetario agregado que el consumidor ha gastado. Estos datos luego se introducen en un *cluster analysis*, que permite identificar a clientes según su frecuencia de compra y su valor adquisitivo, diferenciando así entre clientes nuevos, clientes de baja frecuencia y clientes VIP o fidelizados a la compañía. De esta forma se puede tratar con ellos según los objetivos para el segmento al que pertenecen, ofertando por ejemplo, mejores promociones para aquellos clientes menos fidelizados.

Otra estrategia utilizada es la de segmentar a los clientes según su capacidad de pago. De esta forma se puede mejorar la accesibilidad de familias de renta media-baja y baja a productos financieros de bajo coste adecuados a su perfil de ingresos, capacidad de pago y nivel de aversión al riesgo (García Montalvo, 2014). Esto permite a las entidades facilitar el acceso a servicios financieros a grupos de la población que hoy en día no utilizan los mismos.

Conocer el valor adquisitivo, la edad, datos geográficos, de empleo, el estado familiar y mucha más información de los clientes, permiten tratar con estos de la forma que mejor se ajuste a sus necesidades y llevar así un *targeting* avanzado para ofrecerle a ese segmento productos que van a necesitar y utilizar.

¹² Análisis de grupos o segmentos.

¹³ Actualidad, frecuencia, y monetaria.

2.1.3 Marketing y *targeting* avanzado para empresas

Estos mismos métodos para la segmentación de clientes de la entidad financiera, se pueden utilizar para segmentar y aplicarlo al marketing de las empresas, sobre todo aquellas pymes clientes de la entidad. Se trata de un nuevo servicio que se puede ofrecer a estas compañías, con el fin de aumentar la **eficiencia** y la tasa de conversión de sus compañías de marketing, aprovechándose de los datos que la entidad financiera tenga de sus clientes. La información obtenida de los clientes, organizada y analizada por la tecnología *big data*, permite a las entidades desarrollar estrategias de marketing que verdaderamente atraen a sus clientes, dependiendo de sus características y las del segmento en el que se encuentran.

Otro ejemplo de esto es Commerce 360¹⁴, una herramienta desarrollada en 2016 orientada a la digitalización de pymes.

Mediante la misma se permite que estas empresas puedan “conocer datos de sus clientes, la evolución de las ventas, comportamiento del sector de cara a optimizar recursos, trazar estrategias más eficientes y poner en marcha acciones de marketing. Se convierte los datos anónimos y agregados de compras de clientes con tarjetas en TPV físico, en información útil para los negocios de sus clientes” (Gutiérrez, 2016)

Esta misma tecnología puede ser utilizada por la propia entidad para sus campañas de marketing, aplicando la enorme cantidad de datos sobre el mercado y de sus actuales clientes a sus necesidades personalizadas.

Esta aplicación del *big data* es sobre todo importante en las empresas en el sector financiero debido a la importancia en este sector del *cross-selling* o **venta cruzada**, mediante la cual las entidades financieras tratan de vender productos complementarios a sus clientes actuales para tratar de sacarles una mayor rentabilidad. La información que proporciona el *big data*, sobre como los clientes de la entidad están usando sus productos, permite a la empresa conocer que otros servicios complementarios podría necesitar y, de esta manera, aplicar esta información a sus campañas de marketing.

¹⁴ <https://www.commerce360.es>

2.1.3.1 Análisis de la información del *weblog*

Un aspecto específico donde se puede ver el uso de esta nueva tecnología en cuanto al marketing de las entidades financieras es en el análisis de la información de sus páginas web, donde ofrecen y venden sus productos.

Aún hay mucha información inutilizada por las entidades financieras que les pueden ayudar a entender mejor a sus clientes. Las páginas web de estas entidades generan enormes cantidades de datos a partir de los **clicks** y las **búsquedas** de sus clientes. De esta forma se puede conocer qué productos visitan y qué buscan con mayor frecuencia los consumidores, para así conocer mejor sus necesidades e intereses.

Por sí solo esta información puede ser difícil de analizar y poco fiable. La tecnología *big data* es capaz de almacenar y analizar esta información y combinarla con los datos personales del cliente y su historial de productos para entender mejor sus intereses y necesidades.

2.1.4 Gestión de multicanalidad

El *big data* también puede ayudar a las empresas a gestionar de manera más eficiente su multicanalidad. De esta forma se obtiene información de sus interacciones con clientes y proporciona a estos una mejor experiencia con la empresa, así como con los productos y servicios que ofrece.

La multicanalidad es, por tanto, el conjunto de canales de comunicación que establece la empresa con el cliente y viceversa. Las mayores entidades financieras de España gestionan a diario miles de interacciones con sus clientes. Estas empresas tienen la oportunidad de agregar enormes cantidades de datos de sus interacciones con los clientes, aprovechando al máximo su potencial y teniendo siempre por delante los criterios de eficacia y rentabilidad.

Sin embargo, también supone para las empresas el reto de tratar de combinar los datos obtenidos de los diferentes canales, de forma que estén perfectamente sincronizados e integrados, con el fin de que el diálogo con el cliente sea **coherente** y así mejorar su experiencia y relación con la entidad. Es aquí donde se debe destacar la característica de “veracidad” de la tecnología *big data*, lo cual otorga la capacidad de combinar datos de

orígenes distintos y formatos no homogéneos, y así aumentar la capacidad de análisis de las entidades.

Cabe destacar en este apartado una vez más el estudio desarrollado por CDC software, que revela que los bancos y cajas españoles pierden un 12 % de sus clientes anualmente debido, en gran medida, a la insatisfacción del cliente (Rojas, 2011). Con el desarrollo de un servicio multicanal más eficiente, y de mejor **servicio al cliente**, estas entidades serán capaces de conocer mejor los motivos por los que sus clientes abandonan la entidad, y posiblemente prevenirlo antes de que ocurra.

En este ámbito en España, se puede resaltar el proyecto BBVA Invest, que ha desarrollado su tecnología multicanal a través del *big data*, para así crear un nuevo modelo de asesoramiento a la hora de invertir y utilizar la información obtenida por esta tecnología para adaptarlo al cliente. Este sistema es capaz de mantenerse coherente en todas las plataformas de la entidad gracias a la tecnología *big data* y a su capacidad de analizar y organizar datos de fuentes muy diversas (Alameda, 2017).

La información recopilada de los clientes para este tipo de sistema multicanal permite conocer su perfil de riesgo y su horizonte temporal de inversión, lo cual permite asesorar al cliente de forma personalizada, coherente y a través de todas las plataformas que ofrecen. Este mismo concepto puede ser aplicado a cualquier institución en el área de *financial services* con el fin de que los productos sean más accesibles al cliente y, de esta manera, su interacción con la entidad más individualizada y coherente.

Por otro lado, en el Banco Santander, a lo largo de 2014, ya empezó un avanzado desarrollo de un modelo de distribución multicanal a partir de numerosas nuevas aplicaciones que facilitan la obtención de información del cliente. En España en especial se ve un enfoque hacia la **accesibilidad** y, este banco quiere utilizar su tecnología multicanal, potenciada por la tecnología *big data*, para ofrecerle los mismos servicios de una oficina a sus **clientes digitales** (Santander annual report, 2014).

La mejora de los de la multicanalidad y el acceso para el cliente a partir de la tecnología *big data* favorece la creación de clientes digitales, es decir, los clientes contratan e interactúan con su servicio financiero a través de las diversas plataformas digitales que estas empresas ofrecen.

Estos clientes digitales son más fáciles de atraer, tiene acceso a más productos (lo cual incrementa la venta) y, al mismo tiempo, producen menos costes en cuanto a *overhead costs*¹⁵ de oficinas y personal para la entidad, por lo que el desarrollo de estas tecnologías centradas en el *big data* son claramente ventajas diferenciales para estas empresas.

2.1.5 Asesoramiento financiero

El asesoramiento financiero era un ámbito de la gestión de *big data* en los servicios financieros que hasta hace bien poco, no tenía mucha relevancia en las grandes entidades financieras, y mucho menos en las pequeñas *start-ups* que comenzaban a operar en esta área. Sin embargo, poco a poco esta aplicación del *big data* va cobrando más importancia como servicio para las grandes empresas, debido al desarrollo de la segmentación de los clientes en las entidades financieras, la importancia de la fidelización de los clientes y la enorme cantidad de datos internos y externos que estas empresas tienen de sus clientes.

Un ejemplo de esto es la empresa Smartpig¹⁶, que ya está empezando a trabajar con algunas entidades financieras españolas, utilizando información sobre los gastos categorizados del cliente y la predicción de *cash flow*, permitiendo así al cliente fijar unos **objetivos de ahorro**, así como contribuir a esos objetivos utilizando un tipo de interés atractivo y premios para el cliente que alcance sus objetivos. Simple.com¹⁷ es otro ejemplo claro. Se trata de un banco digital que ya tiene socios financieros en España, y que permite a sus clientes la posibilidad de fijar metas de ahorro y organizar sus gastos en categorías, teniendo en cuenta la **predicción de gastos** que aún no se han producido (García Montalvo, 2014).

Está práctica en las grandes entidades financieras ya se lleva a cabo a través de las oficinas de las entidades. Sin embargo no hay un sistema **automatizado**, inteligente y eficiente como el que la tecnología *big data* puede ofrecer. Este servicio ayuda a la fidelización del consumidor y a conseguir su confianza, además de promocionar entre sus clientes la mentalidad de ahorrar para luego invertir estos ahorros en nuevos productos.

¹⁵ Gastos o costes generales.

¹⁶ <https://www.smart-pig.com>

¹⁷ <https://www.simple.com>

2.2 Gestión de riesgos

El sector financiero debe tener siempre en mente los riesgos que pueden incurrir en cualquier tipo de transacción.

Como bien explica Julián García Barbosa en su artículo, “al aceptar depósitos por un lado y ofrecer créditos por otro, los bancos y cajas de ahorro incurren en riesgos y aquí precisamente radica el beneficio del negocio bancario, por lo que banca y riesgo van de la mano” (García Barbosa, 2013) y esto también es aplicable a todo tipo de entidad financiera que lleve a cabo cualquier tipo de inversión. El *big data* puede ayudar a estas entidades a **optimizar la detección** de riesgos implícitos en su negocio y, a menor riesgo, aumenta la rentabilidad de sus inversiones.

2.2.1 Riesgo de crédito

El principal riesgo que incurren las entidades financieras, y sobre todo los bancos, es el riesgo de crédito. Y es aquí donde se ve principalmente como se está aplicando esta tecnología y como se puede usar el *big data* para analizar este riesgo de manera más precisa.

Al hablar de riesgo de crédito se hace referencia al que puede incurrir una entidad, sobre todo un banco, al conceder un préstamo, aval o tarjeta de crédito a sus clientes.

Como se ha explicado anteriormente, la tecnología *big data* es capaz de obtener y organizar enormes cantidades de datos de los clientes de la compañía, así como combinar los diferentes recursos de información para tener un conocimiento integral del cliente. Esto permite a los bancos desarrollar modelos de *rating* y *scoring*¹⁸ mucho más eficientes con esta tecnología, lo cual minimizaría el riesgo de crédito que la entidad pueda incurrir y dará una imagen más precisa de este riesgo.

A su vez, ayudará con la gestión de **morosidad**, ya que la entidad será capaz de identificar mucho antes si un crédito corre peligro de entrar en mora, pudiendo llevar a cabo las acciones necesarias para minimizar las consecuencias y la forma de renegociar la deuda.

¹⁸ Métodos utilizados por las entidades financieras para calificar la solvencia de particulares y/o empresas que solicitan financiación bancaria.

Hoy en día el análisis de este riesgo se lleva a cabo sintetizando un valor llamado *credit score*, que indica la probabilidad de impago de un crédito. Este valor representa la calidad crediticia de un cliente, como una estimación de la probabilidad de que dicha persona retorne al crédito. Para desarrollar este valor los bancos se basan en datos internos, como las relaciones previas con los clientes, transacciones pasadas, información disponible sobre renta y riqueza, etc. Sin embargo, estos datos no garantizan el pago del crédito y hacen que sea difícil analizar el riesgo de clientes con un historial crediticio breve o sin historial. Por ello hace falta renovar esta visión de solvencia crediticia, aplicando la tecnología *big data* y todos los recursos que ofrece. Esto ha hecho que cada vez gane más relevancia el llamado ***credit score social*** (Óskarsdótti et al., 2018), que también considera la reputación y el estatus social *online* del cliente y de sus contactos, a la hora de analizar el riesgo de crédito del mismo.

Debido a este nuevo sistema de análisis de riesgo, están apareciendo nuevas empresas con su propio sistema de *credit score* que pueden servir de modelo para las grandes entidades financieras. Por ejemplo, la *start-up* alemana Kreditech¹⁹, que concede pequeños créditos en Alemania, Polonia y España, y basa sus estimaciones en información obtenida de Facebook (Vasagar, 2016). En este caso los demandantes cuyos amigos hayan pagado su crédito, tienen mejores trabajos, o viven en mejores barrios, tendrán más facilidad para obtener un crédito. Según señala la empresa, una solicitud tarda menos de 8 segundos en ser analizada y tiene menos de 10 % de impagos. La *start-up* Lenddo²⁰ es otra empresa que utiliza las redes sociales para estimar el “capital social online” de sus clientes, creando un valor entre 0 y 1000 basado en el Facebook de su cliente, sus amigos, las características de estos, su nivel educativo, su trabajo, y el historial crediticio de sus amigos (Wei et al., 2016).

Otras empresas utilizan la red social LinkedIn para crear el *credit score* de sus clientes. Neo Finance se especializa justo en este modelo de análisis de riesgo, creando valores de riesgo para solicitantes jóvenes con poco o ningún historial crediticio, que quieren adquirir un vehículo, pero para los cuales conseguir un crédito por canales habituales sería imposible o tendrían unos intereses muy altos (García Montalvo, 2014). Con su tecnología *big data*, y aplicando los datos obtenidos de LinkedIn, la empresa es capaz de

¹⁹ <https://www.kreditech.com>

²⁰ <https://www.lenddo.com>

predecir la estabilidad de empleo y los ingresos futuros. También es capaz de predecir la rapidez con la que volvería a conseguir un trabajo, utilizando los contactos del cliente con persona de otras empresas. Todo esto con el objetivo final de **predecir la estabilidad laboral** del cliente y ofrecerle un crédito que se ajuste en mejor medida a sus características.

La mayoría de la información de este modelo se obtiene de Facebook, Twitter y LinkedIn, lo cual supone un problema para las grandes entidades financieras, ya que estos datos suelen contener “mucho ruido y poca señal”, a diferencia de los datos que utilizan en sus modelos hoy en día que, por lo general, son más objetivos y de alta calidad. Sin embargo, sería lógico completar su modelo actual con las capacidades predictivas del *big data*.

Khandani et al. (2010) demuestra cómo se pueden mejorar los modelos de predicción del riesgo de impago utilizando algoritmos de *machine learning*²¹ y tecnología *big data*, en particular, combinando la información sobre las transacciones de los clientes y sus *credit scores* generados por agencias de calificación de la calidad crediticia de los consumidores. La aplicación de este modelo consigue una mejora muy significativa en el análisis de este tipo de riesgo. La mejora en la predicción del impago alcanza un 85 % con un ahorro entre el 6 % y el 25 % de las pérdidas totales (Khandani et al., 2010).

2.2.2 Otros riesgos

De la misma manera que se ha visto cómo afecta esta tecnología al análisis del riesgo de crédito, los mismos métodos pueden ser utilizados para analizar otro tipo de riesgos en los que las entidades financieras también incurren. Estos son, por ejemplo, el **riesgo de liquidez**, que hace referencia a la posible incapacidad por parte de la entidad de atender a los diferentes compromisos de pago, o el **riesgo de mercado**, que analiza el riesgo de posibles variaciones en el precio y posición de algún activo de la entidad por la fluctuación de los diferentes mercados.

Al poder analizar datos estructurados y no estructurados a tiempo real y de diversas bases de datos, esta tecnología es capaz de hacer predicciones a futuro que permitan a la

²¹ Rama de la inteligencia artificial. Se basa en la idea de que los sistemas pueden aprender de datos, identificar patrones y tomar decisiones con mínima intervención humana.

empresa adaptarse a posibles cambios de mercado o a una posible falta de liquidez, dándoles así una ventaja estratégica frente a la competencia.

2.3 Gestión de fraudes

En el sector financiero, los fraudes e irregularidades representan uno de los mayores riesgos para la rentabilidad de este tipo de entidades.

Con la tecnología *big data*, reuniendo y asociando los datos de clientes, el mercado y de la propia entidad, estas empresas financieras podrán implementar medidas para la detección de este tipo de actividades fraudulentas (internas y externas) en tiempo real.

Esto permitirá a las empresas en el sector financiero a acabar con uno de los mayores motivos de pérdidas económicas de la organización.

En su análisis de este tipo de aplicación, Barbosa (2013) distingue dos tipos de fraudes a los que se puede aplicar la tecnología *big data* con unos resultados muy positivos:

2.3.1 Detección de fraude en los medios de pago

Cada vez aumenta más el número de estafas que se llevan a cabo mediante medios de pago, y los estafadores cada día tienen a su disposición tecnologías mucho más refinadas para llevar a cabo este tipo de fraudes.

El *big data* permite a las entidades detectar patrones de **comportamiento fraudulento** mediante el análisis histórico de datos en determinadas operaciones llevada a cabo por los clientes. Es importante destacar aquí la V de “velocidad” del *big data*. La autorización de una tarjeta de crédito, o de cualquier transacción, se debe producir de forma casi **inmediata** y, por tanto, es necesario optimizar la detección de operaciones sospechosas. Para llevar esto a cabo las entidades financieras utilizan datos sobre empleados, aplicaciones, fallecidos, encarcelados, listas negras, IRS, etc.

Sin embargo, las entidades financieras aún utilizan una pequeña cantidad de datos que se podrían utilizar para predecir y evitar este tipo de fraude. Por ejemplo, se estima que VISA ha utilizado tradicionalmente solo el 2 % de la información transaccional

(Rosenbush, 2013). La información que utilizan ha ido incrementando a medida que se aplican sistemas de *big data* para facilitar su almacenamiento y análisis. En 2005 utilizaban solamente 40 variables y un solo modelo. Actualmente consideran más de 500 variables, y se prueban 16 modelos diferentes que cubren diferentes mercados y regiones geográficas. La tecnología *big data*, proporcionada por la empresa de software Hadoop, hace posible la actualización rápida de los modelos y datos. Permite que se pueda añadir hasta una variable nueva en una sola hora, cuando en el pasado eran necesarios tres o cuatro días.

Y ésta es una aplicación que ya está en práctica a través de avisos de fraude automatizados, tanto a clientes como a la propia entidad, práctica que ha demostrado ser muy eficiente para las empresas y para sus clientes a la hora de evitar acciones fraudulentas. Sin embargo, este método aún no es del todo eficiente, ya que los estafadores mejoran su tecnología cada día para hacer frente a los nuevos sistemas anti-fraude, generándose aún muchos falsos positivos que afectan negativamente a la satisfacción de cliente y a la imagen de la entidad frente al público.

Hoy en día las entidades financieras utilizan cuatro métodos básicos para detectar fraudes de tarjeta (García Montalvo, 2014). En el primer método los algoritmos de la entidad utilizan desajustes internos directos entre fuentes de datos para detectar el fraude. Este es el método más extendido y utilizado por los bancos. Es, por ejemplo, el uso de una tarjeta en dos países distintos con poco tiempo entre cada transacción, o la falta de coincidencia entre el domicilio de facturación y el envío de la compra de un determinado producto. Estos desajustes pueden provocar la cancelación de la tarjeta. El segundo método se basa en la **detección de outliers**²² o patrones desconocidos, como la inconsistencia en la utilización de la tarjeta o un volumen anormal de facturación. El tercer procedimiento utiliza análisis predictivo en búsqueda de patrones complejos, como buscar transacciones pequeñas con tarjeta en los mismos establecimientos donde se han realizado transacciones previas, lo cual puede indicar que se ha hecho una copia de la tarjeta. Por último, el cuarto método y el menos utilizado por entidades financieras es la utilización de redes sociales y otros datos *online* para buscar patrones de conexión asociativa, como relaciones con conocidos defraudadores o manipulación de identidad (García Montalvo, 2014).

²² Valor atípico.

Es necesario la utilización de la tecnología big data para combinar la información y los resultados de estos cuatro métodos para desarrollar un **sistema híbrido** que se adapte a las nuevas tecnologías, que se ajuste en tiempo real a los nuevos sistemas de los defraudadores y que sea más eficiente para el cliente y evite falsos positivos. Este sistema híbrido, no sólo aumenta la rentabilidad de la compañía evitando las consecuentes pérdidas por la estafa, sino que mejora la imagen de la entidad de cara al cliente.

2.3.2 Prevención de blanqueo de capitales y financiación del terrorismo

El blanqueo de dinero es un proceso mediante el cual redes criminales, incluidos grupos terroristas, pueden distanciar el dinero de su origen sin levantar sospechas y así utilizarlo para financiar sus actividades delictivas.

El *big data* permite a las entidades financieras monitorizar todas aquellas operaciones que la tecnología considere sospechosa, en base a seguir determinados **patrones**. Por ejemplo, cuando un cliente realiza múltiples transferencias de pequeñas cantidades de dinero y luego deposita estas cantidades en múltiples cuentas distintas. La tecnología *big data* es capaz de rastrear estos movimientos y recurrir a datos externos de otras fuentes (redes sociales, bases de datos de seguridad nacional o internacional, etc.) y, gracias al principio de variedad de esta tecnología, puede combinar los datos de estos diferentes orígenes para obtener información sobre el individuo que está realizando la transacción.

Los sistemas actuales para predecir y evitar este tipo de fraude utilizan aún poca información de toda la que está disponible. En la mayoría de casos se limitan a comparar listas de PRP (personas con responsabilidad pública) o SDN (ciudadanos de otros países especialmente susceptibles) o cuentas sospechosas. En la práctica generan muchos falsos positivos y requiere enorme supervisión humana. Este proceso puede ser automatizado y mucho más eficiente con el uso de *big data* y la **combinación** de estos datos con otros indicadores de redes sociales u otros indicios que puedan generarse en el tráfico en internet.

Esta aplicación no está enfocada directamente a la mejora del servicio al cliente, pero de forma indirecta sí es capaz de mejorar la imagen de la entidad y, posiblemente, de mejorar la confianza que estos tienen en la organización.

3. Principales barreras para su aplicación

La introducción de la tecnología *big data* supone enormes cambios para estas grandes entidades, lo cual supone diversas barreras para su implementación y para la obtención de rentabilidad de estos nuevos servicios. A fin de poder introducir esta nueva tecnología en las entidades financieras españolas, las empresas deberán hacer frente a estas barreras y adaptarse a las restricciones legales.

A continuación, analizaré cuales son estas barreras que están impidiendo o limitando la introducción de la tecnología *big data* y qué pueden hacer las entidades financieras para superarlas o adaptarse a ellas.

3.1 Protección de datos

3.1.1 Marco jurídico de la protección de datos

La capacidad de la tecnología actual permite a las grandes entidades de España recopilar una enorme cantidad de datos de sus clientes, la mayoría de veces sin que estos sean completamente conscientes de la información que se está extrayendo sobre ellos mismos a través de las redes sociales, el uso de cookies, las transacciones bancarias, su localización GPS, ... Por ello, el gobierno debe (a nivel nacional y europeo) garantizar el derecho a la vida privada y a la protección de determinados **datos personales**.

Los tribunales españoles definen este derecho, diciendo que “consiste en un poder de disposición y de control sobre los datos personales que faculta a la persona para decidir cuáles de estos datos se pueden proporcionar a un tercero, sea el Estado o un particular, o cuáles puede este tercero recabar, permitiendo también al individuo saber quién posee esos datos personales y para qué, pudiendo oponerse a esa posesión o su uso. Su carácter de derecho fundamental le otorga determinadas características, como la de ser irrenunciable y el hecho de prevalecer sobre otros derechos no fundamentales” (STC 292/2000, de 30 de noviembre de 2000).

En Europa, la protección de datos es un derecho fundamental recogido en el artículo 16 de Tratado de Funcionamiento de la Unión Europea y el artículo 8 de la Carta de los Derechos Fundamentales de la Unión Europea. Y el 24 de octubre de 1995 el Parlamento

y el Consejo Europeo aprobaron la Directiva 95/46/CE, o la Directiva de Protección de Datos, que pretende regular la libre circulación de datos personales en la Unión Europea (Gil, 2015).

En España, este derecho queda recogido en la Ley Orgánica 15/1999 de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD), y su Reglamento (Reglamento de Protección de Datos), aprobado mediante Real Decreto 1720/2007, de 21 de diciembre (LOPD, 1999).

Según establece la LOPD, el objetivo de esta ley es “garantizar y proteger, en lo que concierne al tratamiento de los datos personales, las libertades públicas y los derechos fundamentales de las personas físicas, y especialmente de su intimidad personal y familiar” (LOPD, 1999).

La aparición de las nuevas tecnologías y, sobre todo del uso de *big data* por parte de las grandes empresas tecnológicas, ha hecho que la Unión Europea se replantee y modifique su Directiva de Protección de Datos al haber sido esta publicada en 1995. Por ello se desarrolló el Reglamento General de Protección de Datos (RGPD), que entró en vigor el 25 de mayo de 2016, pero no fue aplicado hasta el 25 de mayo de 2018, permitiendo así a todo tipo de empresas, organismos o instituciones al que aplique este reglamento, 2 años para adaptarse a sus normas (Guía del Reglamento General de Protección de Datos, 2018). En España, el Reglamento también dejó obsoleta la Ley Orgánica de Protección de Datos de 1999, que finalmente fue sustituida por la Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales, la cual entró en vigor el 7 de diciembre de 2018 y a la que las empresas españolas se están empezando a adaptar (LOPD y garantía de los derechos digitales, 2018).

La nueva normativa mantiene las mismas bases e ideas que la anterior, pero queda modificada en ciertos aspectos para adaptarse a las nuevas tecnologías. Por ejemplo, el nuevo reglamento otorga a los clientes de estas entidades el **derecho de acceso** y el **derecho de olvido**, o derecho de supresión de los datos personales, que garantiza al usuario el derecho “al borrado de sus datos personales en el entorno online” (Guía del Reglamento General de Protección de Datos, 2018). Sobre todo innova respecto a la antigua normativa en la introducción del “**principio de la responsabilidad proactiva**” y “**el enfoque de riesgo**”. El principio de la responsabilidad proactiva exige “una actitud

consciente, diligente y proactiva por parte de las organizaciones frente a todos los tratamientos de datos personales que se lleven a cabo” (Guía del Reglamento General de Protección de Datos, 2018), es decir, exige que las organizaciones declaren que datos tratan y con qué finalidad. Por otra parte, “el enfoque de riesgo” se refiere a que las normas de la RGPD se adaptaran a las características de las organizaciones, dependiendo del nivel de riesgo para los derechos y libertades de los ciudadanos de la Unión Europea.

3.1.2 Riesgos para el *big data*

La tecnología *big data* y su gestión presenta un riesgo para la normativa de protección de datos a nivel nacional y europeo. Muchos de los procesos necesarios para llevar a cabo la obtención, almacenamiento y análisis que realiza el *big data*, presentan situaciones dudosas para la normativa actual. A continuación, se llevará a cabo un análisis de algunos de estos riesgos, sobre todo aquellos destacados y recientemente incluidos en la reforma de la normativa de protección de datos, en el Reglamento General de Protección de Datos y en la LOPD. Además, se usarán algunas de las amenazas que la autora Elena Gil (2015) destaca en su proyecto titulado “Big data, privacidad y protección de datos”.

3.1.2.1 Anonimización y el riesgo de reidentificación

La solución utilizada por las empresas que utilizan *big data* y que recogen enormes cantidades de datos para hacer frente a la normativa de protección de datos, ha sido la anonimización de los datos. Este proceso consiste en eliminar todo tipo de **identificadores personales** de las bases de datos, de modo que los individuos no son identificables (Gil, 2015). De esta forma los datos dejan de ser personales, y la normativa de protección de datos no se aplica.

Sin embargo, hoy en día la anonimización tiene limitaciones. El uso de *big data* produce un aumento de la cantidad de información y la capacidad de analizarlo, por lo que es cada vez más fácil la **reidentificación** del individuo, incluso después del proceso de anonimización de los datos.

Hoy en día las autoridades expertas en esta materia ya hablan de **pseudonimización** de datos, ya que ya se considera que la completa anonimización es imposible con la

tecnología actual y su capacidad de analizar y comparar datos (Muineló, 2018).

Actualmente no existe regulación a nivel nacional o europeo de los datos anónimos, aunque si los datos son considerados personales, debido a la alta probabilidad de identificación, serían regulados por la normativa actual.

En teoría, la anonimización absoluta no quedaría regulada por la normativa actual, siendo esta la probabilidad cero de reidentificación, o si existe esta probabilidad, que su riesgo sea insignificante (Gil, 2015). En la práctica esto es imposible. Según la Directiva de Protección de Datos, es necesario “considerar el conjunto de los medios que puedan ser razonablemente utilizados por el responsable del tratamiento o por cualquier otra persona, para identificar a dicha persona” (Directiva 95/46/CE del Parlamento Europeo y del Consejo, 1995).

Por lo tanto, al no existir la anonimización absoluta, las empresas deben reducir el riesgo de reidentificación al mínimo, de modo que funcionalmente sea cero, a pesar de no serlo estadísticamente.

Hoy en día las entidades financieras ya utilizan métodos de pseudonimización, como la **encriptación** o la *tokenización*²³ (Iglesias, 2015), que consiste en codificar la información personal de los clientes. Sin embargo, con el auge de la tecnología *big data* es imposible saber qué información se debe considerar personal, ya que es capaz de relacionar enormes cantidades de datos y encontrar correlaciones entre datos que igual no se consideran relevantes, pero que pueden otorgar mucha información del cliente.

Un ejemplo de esto es el Caso Netflix (Narayanan et al., 2008). En 2006, la empresa proveedora de suscripciones mensuales para ver películas y series de televisión, hizo público el registro de 500 000 de sus usuarios, sin dar la información personal, para ver si alguien era capaz de mejorar el servicio de recomendación. Un individuo, utilizando la tecnología *big data* para combinar y analizar la base de datos de Netflix y la base de datos pública *Internet Movie Database* (IMB), que recopila datos de películas, fue capaz de conocer información personal y sensible de los usuarios, como sus preferencias políticas o su orientación sexual. La Universidad de Texas, utilizando este mismo sistema, fue capaz de identificar en el 84 % de los casos a los usuarios de la compañía, a partir

²³ Un servicio que añade un nivel adicional de protección a los datos sensibles de la tarjeta de crédito, que se reemplazan con un número generado algorítmicamente llamado “token” (Universal Pay, 2017).

solamente de la calificación de seis de sus películas (Narayanan et al., 2008).

Esto indica la gran capacidad de reidentificación a partir de datos aparentemente insignificantes. La dificultad de anonimización de estos datos y el riesgo que esto supone para el *big data*, ya que ahora ya son más los tipos de datos que pueden ser considerados personales, y que quedarían bajo regulación de la normativa de protección de datos. Todo esto se intensifica al entrar en el sector financiero y considerar datos como registros bancarios, perfiles de riesgo o fraude, etc.

Debido al enorme desarrollo de esta tecnología, es imposible que exista una definición concreta de lo que el Gobierno Español y la Unión Europea consideran datos personales, ya que debe ser neutral para adaptarse al rápido desarrollo de la tecnología. Por ello y, para minimizar el riesgo de reidentificación, es imposible atender a la descripción dada por la Directiva de Protección de Datos y citada anteriormente en el apartado.

Es necesario que las entidades españolas utilicen todos sus recursos para minimizar hasta un nivel aceptable la posibilidad de este riesgo, atendiendo al principio de “*privacy first*”²⁴, presentado por el Profesor Josep Domingo-Ferrer en su libro “*Database Anonymization*”²⁵. Este principio indica que las empresas que trabajan con *big data* y analizan cantidades masivas de datos, han de garantizar primero la protección y privacidad de sus clientes (y aplicar esto a la anonimización de sus datos) sin tener en cuenta utilidad que se esté perdiendo en este proceso (Domingo-Ferrer et al., 2016).

3.1.2.2 Consentimiento informado

El consentimiento del cliente supone una de las bases fundamentales para la reforma llevada a cabo por el Reglamento General de Protección de Datos de 2018. La Directiva de Protección de Datos define el consentimiento como: “toda manifestación de voluntad, libre, específica e informada, mediante la que el interesado consienta el tratamiento de datos personales que le conciernan” (Directiva 95/46/CE del Parlamento Europeo y del Consejo, 1995).

Esto supone un enorme riesgo para el *big data*, ya que, tras la recopilación y el análisis

²⁴ Privacidad primero.

²⁵ Anonimización de las bases de datos.

de los datos, en los que es muy importante la anonimización de los mismos, la entidad financiera debe utilizar esta información y aplicarla a las numerosas funciones que han sido analizadas en la primera parte de este proyecto. Para llevar esto a cabo, la entidad necesita el consentimiento del individuo.

La normativa de protección de datos actualizada ya no valora la “inacción” como forma de dar consentimiento, es decir, el cliente debe manifestar su **voluntad** (Guía del Reglamento General de Protección de Datos, 2018). Por ello, hoy en día se utilizan **políticas de privacidad** digitales, a las cuales el cliente ha de dar su consentimiento. Sin embargo, la gran mayoría de los individuos no leen estos documentos, o no son capaces de entenderlo, debido a la gran cantidad de información que se introduce en ellos y el lenguaje especializado que se utiliza para escribirlos (Gil, 2015).

A pesar de que desde un punto de vista legal esto está permitido, realmente no está habiendo un consentimiento informado por parte del cliente, lo cual puede suponer otro tipo de problemas. Sobre todo de cara a la imagen de la entidad al público. Ya se ha visto esto a nivel internacional en 2018, con el caso Facebook-Cambridge Analytica. Aquí, muchos de los datos que habían sido otorgados a la red social Facebook de manera aparentemente informada, fueron luego utilizados por la empresa Cambridge Analytica para desarrollar publicidad política para las elecciones presidenciales de Estados Unidos de 2016 (Stroud, 2018). Esto provocó el desplome de la acción de Facebook en la bolsa de valores americana y una creciente desconfianza por parte de sus usuarios en cuanto a los métodos de la multinacional y el uso de sus datos privados.

Las entidades financieras de España no tienen la completa confianza de sus clientes (Statista, 2012), como se verá más adelante. Por ello, es importante que para impulsar esta tecnología y mejorar la visión de los clientes sobre la entidad, exista **transparencia** en cuanto al uso de los datos entre las empresas del sector financiero español y los clientes que utilizan sus productos.

En primer lugar, es necesario un cambio del lenguaje utilizado en la política de privacidad de estas empresas a un lenguaje más sencillo y más fácil de comprender, para así lograr un consentimiento realmente informado (Gil, 2015). Muchos expertos abogan por el uso de casillas a modo de resumen de la política de privacidad, con el fin de que este sea más fácil de comprender, sin ser el documento oficial al que se da consentimiento. Sin

embargo, al tratarse de un documento legal oficial y, al hablar de empresas en el sector financiero que llevan a cabo transacciones muy delicadas, es importante no perder el contenido y la precisión. Esto es lo que muchos expertos denominan “**la paradoja de la transparencia**” (Gil, 2015), que explica que es inevitable perder cierto contenido cuando el objetivo es la simplicidad.

Sin embargo, esta solución igual no es necesaria. En su proyecto de investigación titulado “*Privacy, big data and the public good*”²⁶, la autora Helen Nissebaum aboga por una nueva solución, una teoría a la que llama la “**tiranía de la minoría**”, y que muchos expertos conocen también como el principio de “minimización de datos”. El principio explica que las grandes empresas son capaces de obtener la misma cantidad de información de unos pocos clientes que sí dan su consentimiento para que se usen sus datos (Nissebaum et al., 2014). Por lo tanto, no sería necesario obtener el consentimiento informado de todos los clientes.

El analista Alan Mislove quiso demostrar justamente esto, y llevó a cabo un experimento en el que trataba de deducir ciertas características sobre un gran número de individuos a partir de los perfiles de redes sociales de unos pocos de sus amigos. El experimento fue un éxito, y tan solo a partir de los datos de un 20% de usuarios, Alan Mislove fue capaz de recaudar ciertos datos sobre el resto de usuarios con un grado muy elevado de precisión (Mislove et al., 2010.)

Potencialmente, esta segunda solución solucionaría el dilema del consentimiento informado y la “paradoja de la transparencia”, ya que no haría falta una simplificación de la política de privacidad. Simplemente se obtendría la información de una minoría de los clientes que den su consentimiento informado y, a partir de sus datos, conociendo su conexión con otros clientes de la entidad a través de sus perfiles de red social y los datos que la entidad tenga de socios y familiares, la tecnología *big data* sería capaz de inferir los datos del resto de sus clientes. Esto también minimizaría errores, al ser menor la base de datos utilizada, y por lo tanto se minimizaría la probabilidad de caer en errores de azar o de confusión.

Sin embargo, la solución también presenta diferentes inconvenientes que también se han de tener en cuenta. En primer lugar, a pesar de que se haya demostrado que la precisión

²⁶ “Privacidad, *big data* y el bien común”

de los datos inferidos es alta, nunca será tan alta como la información que se puede obtener de los datos de todos los clientes. Además, una vez más se estaría trabajando con datos de redes sociales, los cuales suelen ser de dudosa calidad y con mucho ruido. Finalmente, este principio va en contra del propósito del *big data*, que no es otro que analizar una cantidad enorme de datos.

Por lo tanto, existen numerosas soluciones, pero se han de tener en cuenta los aspectos negativos y los riesgos de cada una, siempre trabajando hacia conseguir un consentimiento realmente informado del cliente. La simplificación del lenguaje utilizado en la política de privacidad parece la mejor opción, ya que se mantiene fiel a los principios del *big data*. Sin embargo, se ha de dar un conocimiento completo, sin perder contenido y precisión en la búsqueda de simplicidad.

3.1.2.3 Riesgo en la toma de decisiones automatizadas

La reforma de la regulación de protección de datos en el ámbito europeo, ya ha comenzado a tener en cuenta este tipo de riesgo en el uso de la tecnología *big data* y ya comienza a limitar su uso (Guía del Reglamento General de Protección de Datos, 2018).

Muchas de las decisiones tomadas hoy en día por las entidades financieras españolas son llevadas a cabo de manera totalmente automática, a partir de la información obtenida del *big data* de estas empresas y a través de algoritmos que llevan a cabo decisiones inmediatas, **sin supervisión humana** (Gil, 2015). Estas decisiones influyen, por ejemplo, en el riesgo crediticio de los clientes o las alertas de fraude en las tarjetas de crédito.

Este es por supuesto el objetivo de la tecnología *big data*, minimizar la intervención humana, facilitando así un servicio rápido y preciso para el cliente. Pero esto mismo se convierte en un riesgo cuando se tiene en cuenta la calidad de los datos utilizados, o el riesgo en caer en errores de azar o de confusión al operar con una base masiva de datos, como se demostrará de forma más profunda en el siguiente apartado.

Sin embargo, aún existen numerosas decisiones sobre los clientes de estas empresas que aún se toman de forma automática y que también pueden perjudicar al cliente, como por ejemplo clasificar a un cliente con un perfil de dudoso cobro por algún error en la interpretación de la relación entre variables.

Uno de los riesgos que destaca la RGPD es la creación de **perfiles** de clientes y usuarios de forma automática, sobre todo el riesgo de discriminación que vendría como consecuencia de esto (Guía del Reglamento General de Protección de Datos, 2018). Como ya se ha explicado en la primera parte de este proyecto de investigación, una de las aplicaciones más importantes que puede dar rentabilidad a las entidades financieras de España, es la segmentación de sus clientes en grupos con individuos que se parezcan entre ellos en determinadas características. La reforma de la normativa de protección de datos no prohíbe esta práctica, pero sí limita su uso. Exige transparencia por parte de las entidades en cuanto a que datos son utilizados y con qué fin son necesarios para segmentar a sus clientes, pero sobre todo prohíbe la elaboración de perfiles que tenga como efecto la **discriminación** (Guía del Reglamento General de Protección de Datos, 2018).

Esto supone un enorme problema para los clientes de estas entidades, ya que si los algoritmos discriminan contra ellos, colocándoles en peores perfiles de forma automática, y basándose en datos de baja calidad e información errónea, podría limitar su acceso de ciertos productos que la entidad financiera ofrece.

Por lo que, hasta que exista un mayor desarrollo de la tecnología actual, es necesario la supervisión humana, sobre todo en decisiones trascendentales para los clientes y la entidad, como indica la Propuesta de Reglamento de Protección de Datos (Guía del Reglamento General de Protección de Datos, 2018).

3.1.2.4 Problema de la propiedad de datos

Uno de los principales dilemas que presenta el uso de este tipo de tecnología está en quién es realmente dueño de los datos utilizados por las entidades, y de la información obtenida por el análisis de estos datos.

El Foro Económico Mundial expone de forma clara su posición en este dilema, y explica que, al ser la información creada por ambos grupos, los clientes que proporcionan los datos y las entidades financieras que los analizan a partir de la tecnología *big data*, el derecho no es exclusivo, sino que es común a ambas partes (Foro Económico Mundial, 2012).

Es por ello que la reforma de la normativa de protección de datos haya otorgado a sus clientes el **derecho de olvido**, como se ha destacado anteriormente, dando a los clientes

la capacidad de eliminar los datos personales que la entidad a la que pertenecen tenga sobre ellos.

Es importante que las entidades que hacen uso de esta tecnología sean conscientes de esto, de que los datos que utilizan no pertenecen completamente a la organización y que por tanto, sin un uso apropiado de los datos y una adecuada transparencia con el cliente, estos datos pueden ser borrados de la base de la entidad si así lo exige el cliente (Guía del Reglamento General de Protección de Datos, 2018).

3.2 Calidad de los datos

La cantidad de datos utilizados con la tecnología *big data* supone un riesgo para las entidades que dependan de esta información, ya no a nivel de capacidad de almacenamiento, sino la capacidad de gestión. Como ya se ha destacado anteriormente, el uso de todo tipo de datos, en especial de redes sociales como Facebook o LinkedIn, supone un riesgo para la tecnología *big data*, ya que por lo general son datos muy subjetivos y con mucho ruido, a diferencia de los datos objetivos y de alta calidad a los que están acostumbrados las grandes empresas.

Pero el problema no sólo está en el uso de este tipo de datos, sino en la gestión de las enormes cantidades de datos, los cuales pueden llevar a malentendidos y a errores, que, para las grandes empresas financieras que llevan a cabo transacciones importantes y costosas, puede suponer un riesgo enorme.

La principal función del *big data* es el análisis de enormes cantidades de datos con el fin de así encontrar patrones entre las variables y con ello poder hacer predicciones a futuro. Para ello es imprescindible encontrar la verdadera relación entre variables, es decir, es importante distinguir entre una simple **correlación** y una **causalidad**.

En las ciencias estadísticas, la correlación es la “correspondencia o relación recíproca entre dos o más variables” (Explorable.com, 2010). Esto quiere decir que dos variables estarán correlacionadas cuando el aumento o disminución en una provoca un cambio recíproco en la otra. Así, si el aumento de un valor lleva al aumento de la otra variable, se habla de una correlación positiva. Por lo contrario, si el valor aumenta, pero la segunda

variable disminuye, se habla de correlación negativa. Y si la variable no es afectada por la variación de la otra, es una correlación cero (Explorable.com, 2010).

Sin embargo esto no siempre significa que lo que hay es una relación de causalidad, es decir, que un evento es consecuencia directa del otro. Que existe una **relación causa-efecto** entre las dos variables.

La tecnología *big data* actual es muy buena encontrando la correlación entre una enorme cantidad de datos, sin embargo no siempre es capaz de encontrar la relación causa-efecto entre las variables, lo cual puede llevar a predicciones erróneas basadas en variables que, sí están correlacionadas, pero entre las cuales no existe causalidad.

Existen dos tipos de errores que esta tecnología puede cometer al no poder distinguir entre correlación y causalidad, el error por azar y el error por confusión.

3.2.1 Error por azar

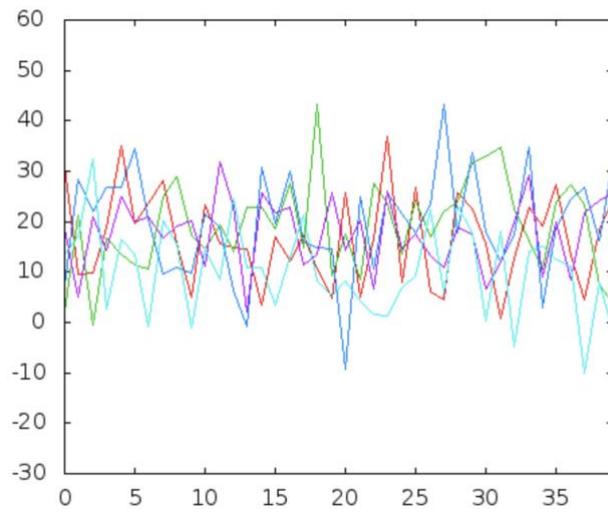
El error por azar, también conocido como el error de los grandes conjuntos de datos significa que, cuanto mayor es el número de datos en el estudio de estas variables, mayor serán las correlaciones que se encontrarán entre estas variables (Gil, 2015). Sin embargo, esta relación estadística no tendrá ningún significado real, al ser una correlación creada como resultado de la enorme cantidad de datos que se han introducido y por tanto no habrá causalidad.

De esta forma, si esta información es interpretada de forma directa, sin ningún estudio posterior, puede llevar a errores y ser un gran riesgo para las entidades financieras que utilizan la tecnología *big data* a la hora de interpretar automáticamente la información de sus clientes.

Esto se puede identificar fácilmente en el experimento del doctor en informática Ricardo Galli (2013), cuyos resultados aparecen en su artículo “Sé cuidadoso con el *big data*”.

En el experimento el autor comienza por introducir cinco variables económicas aleatorias, y como es de esperar no encuentra correlación entre ninguna de ellas.

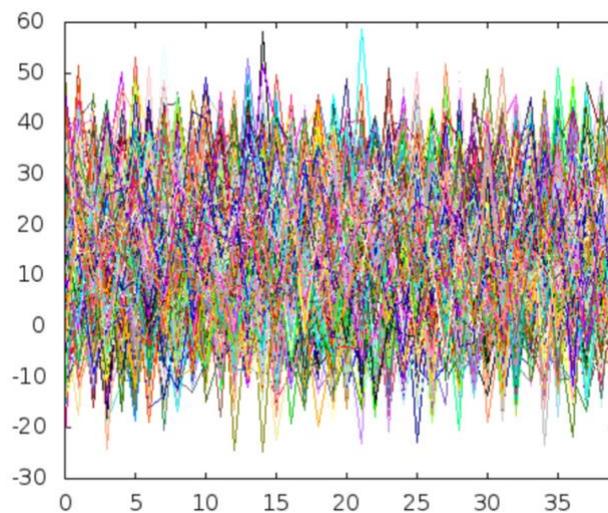
Gráfico I: Variables 1 a 5



Fuente: Galli (2013)

Sin embargo, si se cambia el número de variables, añadiendo mil variables en vez de cinco acercándose así a algo más parecido a lo que lleva a cabo la tecnología *big data*, el autor consigue el siguiente gráfico.

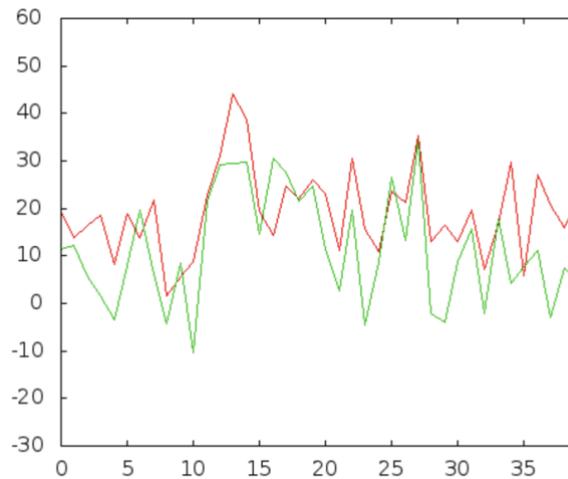
Gráfico II: 1000 variables



Fuente: Galli (2013)

Utilizando el algoritmo Pearson, el autor es capaz de encontrar una correlación entre dos variables de las mil introducidas.

Gráfico III: Variables 30 y 132



Fuente: Galli (2013)

Estas dos variables, la variable 30 y la variable 132, presentan una correlación de Pearson de 0.66, por lo que tienen una alta correlación positiva.

El problema está en que los valores utilizados en el experimento han sido creados automáticamente y de forma artificial con un programa, de forma aleatoria e independiente unos de otros en la misma secuencia y en secuencias diferentes. Por lo que la relación es completamente **aleatoria** y por azar.

Al analizar cantidades masivas de datos, es cada vez más probable encontrar relaciones espurias entre las variables, que en realidad no tienen ninguna relación causa-efecto. Como bien explica el autor, “cuando hay mucha cantidad de datos es muy difícil separar el ruido de la señal” (Galli, 2013).

El riesgo está en que la información obtenida por al análisis de cantidades enormes de datos sea utilizada tal cual, sin un estudio posterior de la causa de un determinado hecho. Para solucionar este problema, la tecnología debe desarrollarse de modo que sea capaz de distinguir entre correlación y causalidad, y ser capaz de encontrar la relación causa-efecto

entre dos o más variables. Llegar hasta esta capacidad de análisis por parte de la tecnología *big data* va a ser un proceso largo y costoso, por lo que hasta que se llegue a este nivel de análisis, la información obtenida deberá ser estudiada de forma controlada, sin una confianza ciega en las capacidades de esta nueva tecnología.

3.2.2 Error por confusión

Este error se produce cuando existen dos variables parecen relacionadas, pero que en realidad son independientes y ambas están relacionadas a una **tercera variable desconocida**, creando así una falsa apariencia de causa-efecto (Gil, 2015).

En su proyecto de investigación titulado “Big data, privacidad y protección de datos”, la autora Elena Gil utiliza el experimento llevado a cabo en 1952 por el matemático polaco J. Neyman para explicar este suceso. En el experimento, J. Neyman, encuentra una relación directamente positiva entre el número de cigüeñas que habitan en los pueblos rurales de Polonia con el número de nacimientos en estos pueblos. Por lo que parece que existe una correlación entre la natalidad de estos pueblos y el número de cigüeñas que habitan en ellos, cuando en realidad esta no es una relación de causa-efecto, sino que existe una tercera variable desconocida que afecta a ambas, en este caso la calidad de las cosechas (Gil, 2015).

Este tipo de error se encuentra fácilmente en la naturaleza, por lo que se intensifica aún más cuando se trata de *big data* y de la enorme cantidad de datos que se analiza. Como se ha representado anteriormente, es muy fácil encontrar relaciones espurias entre variables cuando los datos utilizados son tan grandes, por lo que la posibilidad de caer en este error y no considerar la posibilidad de que exista otra variable que esté afectando a nuestro estudio, es cada vez mayor.

Una vez más, para solucionar este riesgo es necesario un desarrollo de la capacidad de análisis de la tecnología *big data* actual. Sin embargo, éste es un proceso lento y costoso para las grandes entidades. Hasta entonces es necesario un **análisis crítico** de los resultados automáticos obtenidos por esta tecnología y rigor científico en la búsqueda de la causa por la que se produce esta relación.

3.3 Capacidad de la tecnología

Aún se desconoce cuál es la verdadera capacidad analítica de la tecnología *big data* y si existe algún límite. Actualmente, el problema está en la perfección de la tecnología que ya existe y en solucionar los problemas propios del análisis masivo de datos, como son los errores que esta tecnología puede producir (como se ha visto en el apartado anterior). Sin embargo, con vistas a futuro, no se sabe si existe un límite al número de datos que esta tecnología es capaz de almacenar y analizar.

Una encuesta llevada a cabo por la consultora Ernst and Young reconoce esta barrera tecnológica como la principal preocupación del 40,6% de las entidades financieras de España (EY.com, 2015).

Esto es sobre todo visible al analizar la cantidad de datos que están siendo creados y como esta cantidad no deja de crecer. Por ejemplo, desde el comienzo de la historia hasta 2003 los seres humanos han generado 5 exabytes de datos digitales. En 2011 ya se creaban esta misma cantidad de datos en tan solo dos días (Beneyto, 2013).

En cualquier caso, es imposible conocer el límite de la capacidad de análisis de esta tecnología. En su artículo “los límites de la analítica de la información”, el consultor digital Pablo F. Iglesias plantea cuatro etapas en el desarrollo de la capacidad de análisis del *big data*.

En primer lugar, “la pre-interrupción”, que es la situación actual del desarrollo de esta tecnología. Posteriormente viene “la interrupción”, cuando las capacidades predictivas del *big data* permiten que la mayoría de operaciones del sector financiero se desarrollen de forma automática, a partir de algoritmos y de información obtenida de la tecnología *big data*. La tercera etapa la denomina “**el rebasamiento**”, el momento en el que “la recopilación de información evoluciona un ritmo superior que el análisis” (Iglesias, 2012), el límite de la capacidad tecnológica del *big data*. Finalmente, la cuarta etapa es “la síntesis”, cuando se deberá filtrar la información para obtener aquellos datos que sí son de interés para las empresas, limitando así las capacidades del *big data* en el análisis de cantidades masivas de datos.

Es importante que las empresas españolas tengan esto en cuenta, este crecimiento en la creación de datos, para así aplicarlo al desarrollo de su tecnología.

3.4 Falta de conocimiento especializado

Muchas de las grandes empresas españolas, y de las consultoras que investigan esta materia, hablan de la falta de conocimiento especializado y de empleados formados para este tipo de trabajo como una de las barreras fundamentales para el desarrollo del *big data* en el sector financiero de España. El “Informe sobre *Big Data* y *Analytics* en el sector financiero español”, desarrollado por la consultora Ernst and Young en 2015, destaca el crecimiento de la preocupación entre las empresas de este sector por la falta de personal especializado. En un año, el porcentaje de las grandes empresas españolas que consideran que esta es la mayor barrera para el desarrollo de la tecnología *big data* en España, ha subido de 25 % a 47.1 % (EY.com, 2015).

En una entrevista con el periódico *Expansión*, Elena Gil Lizasoain, CEO de LUCA, la unidad global de *big data* en Telefónica, destaca este como uno de los principales problemas que afrontan las empresas españolas. Como bien explica: “son talentos complicados de encontrar en el mercado y sobre todo muy demandados, ya que somos muchas empresas las que estamos intentando incorporarlos” (Pontaza, 2019).

Gil dice que es importante que las empresas que desarrollan la tecnología en esta área estén **involucradas** en la formación de los individuos, para así desarrollar empleados según el perfil de conocimiento que las empresas buscan (Pontaza, 2019).

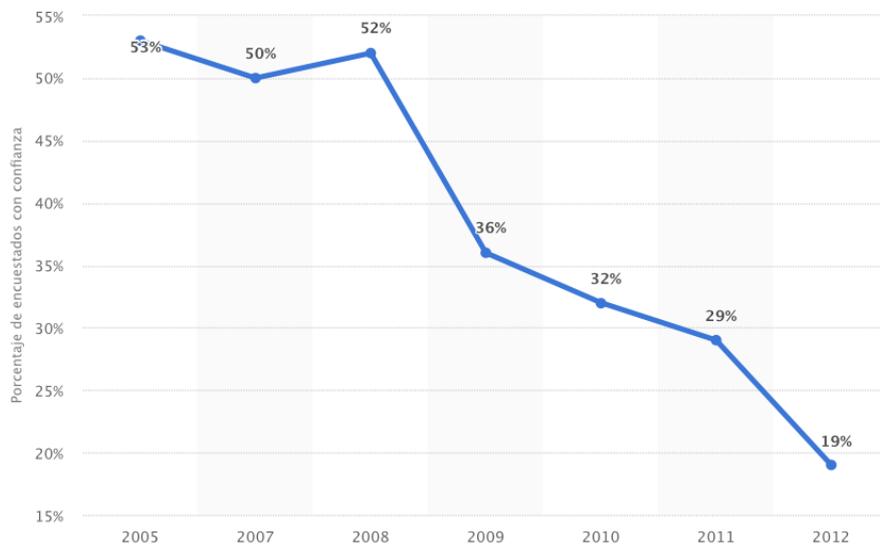
Grandes empresas tecnológicas ya están llevando esto a cabo a nivel internacional, y es importante que se introduzca esta práctica en España. Por ejemplo, la multinacional IBM ya trabaja con universidades e instituciones educativas de Latinoamérica para formar a sus estudiantes y que desarrollen las habilidades asociadas a los requerimientos de la transformación digital que la empresa necesita (Pontaza, 2019).

Si se quiere aumentar el conocimiento especializado en el área *big data* en España, para así ajustarse mejor a la demanda de las grandes entidades financieras, es importante no solo atraer a estudiantes a esta área a través de incentivos salariales y otros beneficios, sino que las grandes empresas españolas trabajen con los centros educativos para desarrollar el perfil de empleado que necesitan.

3.5 Confianza de los clientes (desconfianza en las entidades)

La confianza de los ciudadanos en las instituciones financieras españolas ya no es lo que era, sobre todo en cuanto al sector bancario. Esto se puede ver claramente en la estadística desarrollada por la empresa de investigación y análisis Statista, que refleja el nivel de confianza que la población española en sus instituciones financieras y en sus bancos entre 2005 y 2012.

Gráfico IV: nivel de confianza de la población española en sus instituciones financieras y en sus bancos entre 2005 y 2012



Fuente: Statista (2012)

Esta caída se debe a varios factores. En primer lugar, la **crisis financiera de 2008**, generada principalmente por los grandes bancos, y que, junto al estallido de la burbuja inmobiliaria este mismo año, dejó a España en una gran recesión, de la cual aún se ha recuperado por completo. Esto, sumado a las constantes noticias de corrupción y la falta de transparencia entre estas entidades y sus clientes, hace que la desconfianza de los españoles no deje de crecer.

Para llevar a cabo un buen desarrollo de la tecnología *big data* es necesario una plena confianza en las instituciones. La manipulación y el mal uso de esta tecnología puede

potencialmente aumentar esta desconfianza y limitar aún más su desarrollo. Esto es aún más pertinente cuando se tiene en cuenta los escándalos que el mal uso de esta tecnología está creando, como es el caso Facebook-Cambridge Analytica (Stroud, 2018) que se ha citado anteriormente.

Por ello, es necesario que las grandes entidades financieras españolas se adapten a las limitaciones legales que existen y que quedan recogidas en la Ley Orgánica de Protección de Datos de Carácter Personal y así llevar a cabo un desarrollo responsable de la tecnología *big data*, atendiendo a los riesgos legales y también éticos que existen al obtener, almacenar, y analizar estas enormes cantidades de datos personales.

4. Conclusiones

Aunque todavía en una fase incipiente, ya se comienza a ver el enorme papel que tiene el *big data* en el futuro del sector financiero del país.

A pesar de llegar tarde en España, los resultados prácticos de esta tecnología en las entidades financieras del país no dejan de crecer, siempre de mano de las grandes empresas, las cuales ya invierten millones en el desarrollo de la tecnología y apuestan por su importancia en la mejora de su negocio. Al mismo tiempo, pymes y *start-ups* continúan superando las barreras que supone esta metodología y revolucionando el uso que se le puede dar.

Como se ha planteado a lo largo del presente trabajo, este proyecto no sólo analiza las utilidades, ventajas estratégicas y utilidad que esta tecnología *big data* puede ofrecer, sino que también se desarrolla las barreras que se han de superar para alcanzar el verdadero potencial de la herramienta y los riesgos que acompañan a una confianza ciega en sus capacidades.

Las oportunidades estratégicas del *big data* son reales, y necesarias para el desarrollo de este sector. Comenzando así por el **servicio al cliente**, la función más importante que debe tener el *big data* en estas entidades. Es aquí donde la entidad financiera es capaz de aumentar su rentabilidad respecto a sus clientes, aumentando la fidelización de los mismos y la venta cruzada de productos, a través de un servicio especializado que aumenta la

confianza y satisfacción del cliente que, como se ha analizado en el apartado anterior, requiere una mejora.

En segundo lugar está la **gestión de riesgos**, principalmente de crédito, pero también de liquidez o de mercado. Este representa el mayor peligro para los bancos y cajas de ahorro españolas, al igual que otras entidades en el área de servicio financiero, ya que el beneficio de estas entidades radica en el beneficio que se obtiene de los depósitos y créditos. En el apartado se analiza como la rentabilidad del *big data* en esta aplicación se obtiene de una mayor eficacia y eficiencia en el desarrollo de modelos de *rating* y *scoring*, no solo a partir de un análisis a tiempo real del mercado y la capacidad de pago de los clientes, sino a partir de un aprovechamiento de una mayor fuente de datos.

El trabajo es quizás más radical en esta propuesta, la del uso de un **modelo híbrido** entre las fuentes de datos tradicionales y nuevas fuentes de información no tan “objetivas” (tipo RRSS). Ya existen start-ups que utilizan un modelo similar, obteniendo resultados muy prometedores pero, como se resalta sobre todo en la segunda parte del proyecto, existe un riesgo en el uso de estos datos. La propuesta final ante esta gran variedad de datos, con procedencias distintas y de mayor o menor “fiabilidad”, no consiste en una confianza ciega en la calidad de esta información, sino en un modelo que de preferencia a las fuentes tradicionales de datos de alta calidad, utilizando así este segundo grupo de datos como complemento a la información obtenida, por lo menos hasta que la tecnología se desarrolle de tal forma que sea capaz de distinguir los datos de alta calidad de aquellos con posibilidad de error (“mucho ruido y poca señal”).

Finalmente, cabe destacar la ya bastante desarrollada **gestión de fraudes** a través de esta tecnología, por un lado en los medios de pago y por otro en el blanqueo de capitales.

En el primer caso, hoy es evidente la aportación del *big data* en el aumento de rentabilidad para la entidad, mediante la reducción del enorme gasto directo que siempre ha supuesto el fraude. Históricamente se ha utilizado muy poca información transaccional y modelos únicos para la detección de fraudes específicos. El *big data* permite analizar los datos históricos de las operaciones y el desarrollo de un modelo híbrido que combina información muchísimo más amplia y diversa. En segundo lugar, a pesar de que el blanqueo de capital no sea un gasto directo que incurren estas empresas, sí que afecta a

la imagen de la entidad frente a sus clientes y a la confianza que estos tienen en ella (actualmente unos niveles históricos muy bajos).

La segunda parte del proyecto consiste en el **análisis de los riesgos legales, técnicos, y de mercado** que supone el uso de esta tecnología y al que las entidades financieras se deben adaptar.

Según esta segunda parte, el desarrollo de esta tecnología debe estar enfocado a un mayor desarrollo de las **capacidades analíticas** de los datos, para así obtener resultados de alta calidad, y evitar los errores propios de un análisis masivo de datos. Es decir, la tecnología ha de ser capaz de analizar la relación causa-efecto entre variables, y no simplemente encontrar correlaciones. Es por ello por lo que se recomienda un análisis crítico de los resultados obtenidos por esta tecnología y rigor científico en la búsqueda de la causa por la que se produce la relación.

Desde el punto de vista del mercado, el proyecto resalta la **falta de personal cualificado** que existe en la actualidad en el país. Es necesario que las entidades de España trabajen para fomentar los estudios y desarrollar el perfil de trabajador que requieren.

También desde el punto de vista del mercado habría que resaltar la **desconfianza de los clientes**, un dato de especial importancia en el desarrollo de esta herramienta y sus repercusiones prácticas, ya que es necesaria esta confianza cuando se habla de una tecnología que recopila y utiliza información tan sensible. Por ello, se recomienda transparencia por parte de las entidades, así como un desarrollo responsable, ético, y legal de la tecnología.

Y, en último lugar, es necesario afrontar **el riesgo legal** que supone el propio uso de la tecnología en cuanto a un ataque a la privacidad de las personas. Las normativas de protección datos nacional y europea fueron actualizadas hace menos de un año y este nuevo escenario supone un riesgo para el actual uso del *big data*, por lo que este apartado es de especial relevancia y supone una verdadera revolución muy a tener en cuenta en el uso de la tecnología. Se presentan diversas soluciones para estos problemas, como la minimización de las decisiones automatizadas (sobre todo aquellas de trascendental importancia para los clientes y la entidad), el uso de un lenguaje simplificado para las políticas de privacidad con el fin de obtener un consentimiento realmente informado y,

finalmente, una minimización a niveles aceptables del riesgo de reidentificación, atendiendo siempre primero a la protección y privacidad de los clientes y no a la posible utilidad que se pueda perder.

La carrera por el desarrollo de esta tecnología ya ha comenzado entre las entidades financieras españolas, por un *big data* más eficaz y eficiente, ya que **la información es ahora el recuso de mayor importancia para las empresas**. Se ha destacado en el proyecto la rentabilidad y la ventaja estratégica que esta tecnología puede otorgar a las entidades que consigan hacer uso de su verdadero potencial. Sin embargo, esta rentabilidad se perdería sin un desarrollo responsable de la tecnología. Se ha de atender a las barreras tecnológicas actuales que presenta esta herramienta y como ha de afectar esto a su desarrollo futuro. Además, es importante que el sector financiero español se adapte a las limitaciones legales del *big data*, y que quedan recogidas en la Ley Orgánica de Protección de Datos de Carácter Personal, al igual que sus posibles reformas futuras, para lograr así un avance responsable y ético en estos nuevos sistemas de trabajo.

Bibliografía

Alameda, T. (2017). *Lo que el 'big data' financiero revela y cómo se usa en BBVA*. [online] BBVA NOTICIAS. <https://www.bbva.com/es/big-data-financiero-revela-usa-bbva/>.

Barbosa, J. (2013). *El papel de big data en tres áreas críticas del sector bancario*. [online] A un Clic de las TIC. <https://aunclidelastic.blogthinkbig.com/el-papel-de-big-data-en-tres-areas-criticas-del-sector-bancario/>.

Beneyto, R. (2013). *¿Cuánta Información se Genera y Almacena en el Mundo?*. [online] Documanía 2.0. <https://documania20.wordpress.com/2013/09/16/cuanta-informacion-se-genera-y-almacena-en-el-mundo/>

Datio. (2019). *Datio Big Data*. [online]. <https://www.datio.com>.

Domingo-Ferrer, Josep., et al. *Database Anonymization Privacy Models, Data Utility, and Microaggregation-Based Inter-Model Connections*, 2016.

E. Khandani, A., J. Kim, A. and W. Lo, A. (2010). *Consumer credit-risk models via machine-learning algorithms*. [ebook] Elsevier. <http://alo.mit.edu/wp-content/uploads/2015/06/Household-behaviorConsumer-credit-riskCredit-card-borrowingMachine-learningNonparametric-estimation.pdf>.

Elmer-Dewitt, P. (2011). *Whats 12 petabytes to Apple?*. [online] Fortune. <http://fortune.com/2011/04/07/whats-12-petabytes-to-apple/>.

Explorable.com. (2010). *La Correlación Estadística*. [online] <https://explorable.com/es/la-correlacion-estadistica>

Ey.com. (2015). *EY - Informe sobre Big Data y Analytics en el sector financiero español*. [online] <https://www.ey.com/es/es/home/ey-informe-sobre-big-data-y-analytics-en-el-sector-financiero-espanol#.XMJEUi3MzOQ>

Galli, R. (2013). *Sé cuidadoso con el "Big Data"*. [online] Ricardo Galli, de software. <https://gallir.wordpress.com/2013/05/29/se-cuidadoso-con-el-big-data/>

García Montalvo, J. (2014). *El impacto del "Big data" en los servicios financieros*. [ebook] University Pompeu Fabra. http://www.econ.upf.edu/~montalvo/wp/big_data_banking_v4.pdf.

Gil, E. (2015). *Big data, privacidad y protección de datos*. 19th ed. [ebook] Agencia española de Protección de Datos. <https://www.aepd.es/media/premios/big-data.pdf>

Guía del Reglamento General de Protección de Datos. (2018). [ebook] Agencia Española de Protección de Datos. <https://www.aepd.es/media/guias/guia-rgpd-para-responsables-de-tratamiento.pdf>

Gutiérrez, D. (2016). *Nace Datio, un jointVenture entre BBVA y Stratio para liderar el big data en el sector financiero*. [online] Abra Invest. <http://abra-invest.com/nace-datio-jointventure-bbva-stratio-liderar-big-data-sector-financiero/>.

Hj.tribunalconstitucional.es. (2000). *Resolución: SENTENCIA 292/2000*. [online] <https://hj.tribunalconstitucional.es/es-ES/Resolucion/Show/4276>

Iglesias, P. (2015). *Cómo funciona la tokenización de los pagos y su anonimización*. [online] PabloYglesias.com. <https://www.pabloyglesias.com/tokenizacion-pagos-moviles/>

Iglesias, P. (2016). *Los límites de la analítica de la información*. [online] PabloYglesias.com. <https://www.pabloyglesias.com/los-limites-del-big-data/>

Lane, J., Stodden, V., Bender, S. and Nissenbaum, H. (2014). *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. [ebook] Cambridge University Press. <http://wpressutexas.net/cs378h/images/b/b3/LaneEtAlPrivacyBigDataAndThePublicGood.pdf>

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. (2018). [ebook] Madrid: Agencia Española de Protección de Datos. <https://www.boe.es/boe/dias/2018/12/06/pdfs/BOE-A-2018-16673.pdf>

Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal. (1999). [ebook] Madrid. <https://www.boe.es/buscar/pdf/1999/BOE-A-1999-23750-consolidado.pdf>

Linden, G., Smith, B. and York, J. (2003). *Amazon.com recommendations: item-to-item collaborative filtering*. [ebook] Internet Comput. IEEE. <https://ieeexplore.ieee.org/document/1167344>.

Martinez, J. (2017). *La banca admite su atraso y echa el resto en 'big data'*. [online] El Español. https://www.elespanol.com/economia/empresas/20170329/204480226_0.html.

Mislove, A., Viswanath, B., P. Gummadi, K. and Druschel, P. (2010). *You are who you know: inferring users profiles in online social networks*. [ebook] Nueva York. <https://mislove.org/publications/Inferring-WSDM.pdf>

Muñelo, A. (2018). *El Reglamento Europeo de Protección de Datos ¿Qué es la seudonimización?*. [online] Worldcomplianceassociation.com. <http://www.worldcomplianceassociation.com/1348/articulo-el-reglamento-europeo-de-proteccion-de-datos-que-es-la-seudonimizacion.html>

Narayanan, A. and Shmatikov, V. (2008). *Robust De-anonymization of Large Sparse Datasets*. [ebook] Austin: The University of Texas Austin. https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J. and Baesens, B. (2018). *The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics*. [ebook] University of Southampton. <https://www.sciencedirect.com/science/article/pii/S156849461830560X>.

Pontaza, D. (2019). *Talento, barrera en la implementación de Big Data en México...y el mundo*. [online] Expansión. <https://expansion.mx/tecnologia/2019/02/21/talento-barrera-en-la-implementacion-de-big-data-en-mexico-y-el-mundo>

Predictland.com. (2018). *Big Data: Segmentación de Clientes*. [online]. https://www.predictland.com/big_data_segmentacion_clientes/.

Rethinking Personal Data: Strengthening Trust. (2012). [ebook] Foro Económico Mundial. http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf

Rojas, E. (2011). *Bancos y cajas pierden al año un 12 % de clientes insatisfechos*. [online] MuyComputerPRO. <https://www.muycomputerpro.com/2011/12/27/bancos-cajas-pierden-clientes-por-desconocer-los-motivos-de-su-insatisfaccion>.

Rosenbush, S. (2013). *Visa Says Big Data Identifies Billions of Dollars in Fraud*. [online] WSJ.com. <https://blogs.wsj.com/cio/2013/03/11/visa-says-big-data-identifies-billions-of-dollars-in-fraud/>

Santander annual report. (2014). [ebook] Banco Santander. <https://www.santanderannualreport.com/2014/sites/default/files/estrategia-multicanal.pdf>.

Serrano Acitores, A. and García Martín, L. (2019). *Big Data y Protección de Datos*. [online] Antonioserranoacitores.com. <https://www.antonioserranoacitores.com/big-data-proteccion-datos/>

Statista. (2012). *España: confianza en bancos 2005-2012*. [online] <https://es.statista.com/estadisticas/491478/confianza-publica-en-bancos-espanoles/>

Stroud, C. (2018). *Cambridge Analytica: The Turning Point In The Crisis About Big Data*. [online] Forbes.com. <https://www.forbes.com/sites/courtstroud/2018/04/30/cambridge-analytica-the-turning-point-in-the-crisis-about-big-data/#1c035fd348ec>

UniversalPay. (2017). *¿Que es la tokenización?*. [online] <https://www.universalpay.es/que-es-la-tokenizacion/>

Vasagar, J. (2016). *Kreditech: A credit check by social media* | Financial Times. [online] Ft.com. <https://www.ft.com/content/12dc4cda-ae59-11e5-b955-1a1d298b6250>.

Wei, Y., Yildirim, P., Van den Bulte, C. and Dellarocas, C. (2016). *Credit Scoring with Social Network Data*. Maryland, USA: Institute for Operations Research and the Management Sciences (INFORMS).

Wortham, J. (2013). *A Financial Service for People Fed Up With Banks*. [online] Nytimes.com. <https://www.nytimes.com/2013/01/09/technology/a-financial-service-for-people-fed-up-with-banks.html>.

Zarantech.com. (2016). *The 4 V's of Big Data*. [online] <https://www.zarantech.com/blog/the-4-vs-of-big-data/>.

