CUADERNOS SALMANTINOS DE FILOSOFÍA
Suplemento 1, Vol. 42, 2015, 315-330, ISSN: 2387-0818

BIG DATA: LA ECLOSIÓN DE LOS DATOS

VIDAL ALONSO SECADES
Universidad Pontificia de Salamanca. Salamanca. España
valonsose@upsa.es

Resumen: Medio siglo después de que los ordenadores hayan entrado en nuestras vidas, los datos han empezado a eclosionar hasta tal punto que se puede afirmar que algo nuevo y especial está teniendo lugar. Cuando comienza la era del Big Data, las instituciones y organizaciones empresariales pasan a ser conscientes del alto potencial remanente en los datos que están almacenados en sus archivos. Así, los datos almacenados pasan a ser un activo de la organización y se utilizan para buscar valor en ellos que conduzca hacia una toma de decisiones basada en datos. Sectores como servicios, sanidad, educación o la propia área gubernamental están aplicando estos nuevos métodos de análisis de cara a inferir conocimiento mediante la aplicación del razonamiento deductivo. Pero la transformación de los datos en conocimiento no es una tarea fácil ni directa ya que los conjuntos de datos disponibles se encuentran generalmente desorganizados. Es la era del Big Data que emerge de la eclosión de los datos.

Palabras clave: data mining, privacidad, razonamiento deductivo, tecnología, toma de decisiones.

BIG DATA: THE DAWN OF THE DATA

Abstract: Half a century later that the computers have appeared into our lives, data have begun to dawn in such a way that it can be said that something new and special is taking place. When the era of Big Data begins, institutions and business organizations become aware of the high potential remaining in the data that are stored in their files. Thus, the data stored become an asset of the organization and they are used to find value in them in order to lead to a decision-making based on data. Sectors such as services, health, education or the own government area are applying these new analysis methods in order to infer knowledge through the application of deductive reasoning. But the data transformation into knowledge is not an easy task because the available data sets are usually disorganized. It is the Big Data era which emerges from the data dawn.

Key Words: data mining, decision making, deductive reasoning, privacy, technology.

Cuando acepté la propuesta para participar en este volumen homenaje al Profesor/Filósofo Dr. D. Marceliano Arranz Rodrigo, entré en una profunda reflexión pensando en un contenido que pudiera agradar al profesor homenajeado.

No fue una reflexión sencilla, ya que a lo largo de estos años hemos compartido innumerables temas de debate, diálogo o conversación en nuestras excursiones a la montaña o a la siempre bien intencionada, aunque en ocasiones poco fructífera, tarea de recolección de setas.

Sin embargo, mi reflexión, guiándose por las más recientes conversaciones, me ha llevado a las que considero bases del profesor Arranz. Él, siempre orgulloso de haber sido discípulo del Dr. Bochenski en las clases de la Facultad de Filosofía en la Universidad de Friburgo, intentaba inculcarme alguna de sus clarividentes ideas basadas en el uso de la lógica matemática para mantener el orden de las cosas. Recordaré con especial agrado nuestras charlas paseando por la playa de Xagó, donde el ruido de las olas de la mar y el olor al yodo de la brisa marina estimulaban su intelecto y creatividad hasta límites sorprendentes, ofreciéndole nuevas perspectivas filosóficas de la vida.

Por mi parte, dados mis limitados conocimientos de una materia que me fue impartida durante un sólo semestre a lo largo de mi etapa universitaria, era un simple receptor que carecía de los mínimos argumentos para poner en cuestión sus afirmaciones.

Pero quienes me conocen saben que no soy un oyente pasivo y que disfruto del debate afable que ayuda a confrontar diferentes puntos de vista. Así, ante las deliberaciones del profesor Arranz, contraponía nuevos conceptos de razonamiento deductivo que estaban emergiendo y que me permitían instruir sobre aspectos que Marceliano desconocía. O al menos eso era lo que yo creía.

No hay nada como enfrentar a un buen filósofo con un problema o una nueva labor. Su tenacidad y constancia en el trabajo ha hecho que me haya proporcionado de forma constante lecturas, artículos, libros o teorías que han ido saliendo sobre el tema, hasta el día de hoy, donde estoy convencido que en los fundamentos del nuevo razonamiento deductivo el profesor Arranz está mucho más instruido que servidor.

Una vez expuesta mi reflexión, pretendo ser reiterativo en mis empecinamientos y como última contribución a su etapa académica voy a intentar, una vez más, acercarle algunas de las novedades que el nuevo razonamiento deductivo está aportando a la sociedad y la influencia que en su desarrollo está teniendo.

INTRODUCCIÓN

En 1980, Alvin Toffler definía en su libro "La Tercera Ola" a un analfabeto como:

> Aquel que no sepa dónde ir a buscar la información que requiere en un momento dado para resolver una problemática concreta. La persona formada no lo será a base de conocimientos inamovibles que posea en su mente, sino en función de sus capacidades para conocer lo que precise en cada momento (Toffler, 1980).

No cabe duda que, hoy en día, la formación de una persona pasa por ser consciente de la importancia que los datos están alcanzando dentro la sociedad global en que estamos viviendo. Internet, la tecnología móvil o las redes sociales, están generando tal cantidad de datos que, en numerosas ocasiones, inducen a alcanzar la sensación de compartir nuestra vida privada con organizaciones o empresas totalmente ajenas.

Medio siglo después de que los ordenadores hayan entrado en nuestras vidas, los datos han empezado a eclosionar hasta tal punto que se puede afirmar que algo nuevo y especial está teniendo lugar. Esta eclosión de los datos, entendiendo eclosión dentro de la segunda acepción de la Real Academia Española como "Brote, manifestación, aparición súbita de un movimiento cultural o de otro fenómeno histórico, psicológico, etc.", está dando lugar a una nueva era conocida como la era del Big Data.

Aunque el primer documento académico relativo a Big Data fue publicado en el año 2003 por *Francis X. Diebolt* (Diebold, 2003, 115-122), el término ya fue empleado por primera vez a finales de los noventa por *John Masley* quién impartió una serie de charlas a pequeños grupos acerca del significado que esta nueva era iba a tener. Una era que venía descrita por una rápida expansión del volumen de datos, más allá de lo que la gente pudiese imaginar (Dean, 2014).

Con anterioridad al comienzo de este nuevo periodo las empresas otorgaban un valor casi nulo a los datos que recolectaban en sus transacciones. Cuando comienza la era del Big Data, las instituciones y organizaciones empresariales pasan a ser conscientes del alto potencial remanente en los datos que están almacenados en sus archivos. Esta constatación hace que la tendencia hacia la recolección y almacenamiento de datos cambie y se realice un mayor esfuerzo para mantener y estructurar sus repositorios o almacenes de datos.

De esta forma los datos almacenados pasan a ser un activo de la organización y se utilizan para buscar valor en ellos que conduzca hacia una toma de decisiones basada en los datos que pueden recolectarse de un mayor número de personas, durante un mayor periodo de tiempo, y además, con un coste y un esfuerzo menor al hasta ahora conocido.

En este artículo se pretende mostrar las principales características de Big Data y sus actuales aplicaciones dentro de la sociedad. Como se observará, la transformación de los datos en conocimiento no es una tarea fácil ni directa. Los conjuntos de datos disponibles se encuentran generalmente desorganizados, contienen detalles inútiles y muchas veces el conocimiento encerrado en ellos es incompleto, siendo necesaria una depuración de los mismos que evite la generación de incertidumbre. (Joyanes, 2014)

Esta era llega para quedarse por unos cuantos años. En ella, se ejecutarán actuaciones, basadas en conocimiento generado mediante razonamiento deductivo, que aportarán numerosos cambios y oportunidades a la sociedad. Es la era del Big Data que emerge de la eclosión de los datos.

1. FUNDAMENTOS DEL BIG DATA

El término Big Data genera todavía confusión y suele asociarse a conceptos relacionados con grandes volúmenes de datos, análisis sociológicos, gestión de datos o aplicaciones comerciales. Esta confusión se debe a que no existe una definición rigurosa del término de Big Data.

Inicialmente, es un término que surge ante el gran crecimiento que experimenta el volumen de información a procesar, de manera que los datos a examinar no pueden procesarse con las capacidades de las memorias de los ordenadores, obligando a buscar nuevas herramientas que puedan analizar tal cantidad de datos.

Hoy en día, la idea que prevalece es que se habla de Big Data:

> Cuando los análisis para extraer nuevos hechos, crear nuevas formas de valor, cambiar mercados, organizaciones u observar las relaciones entre la sociedad y los gobiernos, deben hacerse a gran escala y no pueden realizarse a una escala menor (Schönberger; Cukier, 2014).

En una definición más técnica, dada por un referente del sector como IBM, Big Data puede definirse como "la convergencia digital entre los datos estructurados presentes en bases de datos y los datos no estructurados provenientes de las nuevas fuentes de información como redes sociales, dispositivos móviles, sensores RFID, smartphones o sistemas financieros" (IBM, 2013). Esta definición se ve refrendada por la teoría de las 3 V´s del Big Data, que no sólo considera el volumen de datos, sino también la variedad de las fuentes de información y la velocidad de captación de los datos.

Por otro lado, es interesante considerar el concepto desde el punto de vista de las empresas del ámbito de la consultoría. Así, Mckinsey, a través de su *McKinsey Global Institute*, define Big Data como "Conjuntos de datos cuyo tamaño va más allá de las posibilidades de las típicas herramientas de bases de datos para capturar, almacenar, gestionar y almacenar" (Manyika et al., 2011). Por otro lado, la consultora IDC[1] considera que:

> Big Data es una nueva generación de tecnologías, arquitecturas y estrategias diseñadas para capturar y analizar grandes volúmenes de datos provenientes de múltiples fuentes heterogéneas a una alta velocidad con el objeto de extraer valor económico de ellos.

De lo que no hay duda, y se desprende de todas las definiciones existentes, es que la tecnología Big Data permite a las organizaciones capturar y analizar cualquier dato, sea cual sea su procedencia, de cara a aportar conocimiento relevante para efectuar una toma de decisiones basada en la información disponible.

Para analizar toda esta gran cantidad de información disponible se están utilizando técnicas y algoritmos de *Data Mining*, también conocidos como procesos de descubrimiento de conocimiento o procesos KDD[2]. En su forma de trabajo, este proceso KDD explora las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar. Para lograrlo, procesa conjuntos de datos o *datasets* buscando encontrar patrones de comportamiento similares o alcanzar modelos predictivos de conocimiento que puedan deducirse de los datos.

Es necesario reconocer que, ante la gran variedad de fuentes de información, los altos volúmenes de datos o los nuevos modelos de negocios, todavía queda mucho recorrido para aprender como trabajar con Big Data. Pero si una cosa está bien clara, es que los métodos tradicionales de trabajar con los datos no nos conducirán a alcanzar grandes resultados en los análisis de Big Data. Como bien dice West:

> Alguien que recomiende utilizar las mismas viejas herramientas de procesamiento de datos bajo las nuevas circunstancias de captura y almacenamiento de datos, es alguien que está totalmente fuera del análisis de datos (West, 2012).

1 International Data Corporation
2 Knowledge Discovery Databases

1.1. Data mining

Data Mining, o minería de datos, es una tecnología que ayuda a las organizaciones a concentrarse en la información más importante existente en su base de información. A través de las herramientas de *Data Mining* es posible predecir futuras tendencias y comportamientos, permitiendo efectuar una toma de decisiones proactiva y conducida por un conocimiento extraído de la información (*knowledge-driven*) (Guidici, 2003).

El nombre de *Data Mining* deriva de las similitudes entre buscar información valiosa para las empresas en bases de datos gigantes y minar una montaña para encontrar una veta de metales valiosos. Ambos procesos requieren examinar una gran cantidad de material, hasta encontrar exactamente donde residen los valores buscados.

En este proceso los grandes volúmenes de datos conforman el *dataset* que se analizan mediante técnicas y algoritmos con el objetivo de encontrar patrones, tendencias, asociaciones o reglas ocultas que expliquen o predigan el comportamiento de un fenómeno en un contexto determinado (Hernández *et al*., 2004).

Las herramientas de *Data Mining,* basadas en técnicas estadísticas, de inteligencia artificial o de aprendizaje automático, permiten responder a problemas de las empresas que tradicionalmente consumen demasiado tiempo en su resolución. El descubrimiento automatizado de modelos desconocidos o la predicción automatizada de tendencias y comportamientos, como la detección de transacciones fraudulentas en el uso de tarjetas de crédito, pueden ser algunos ejemplos de su aplicación.

Estas aplicaciones otorgan al proceso *Data Mining* de una gran potencialidad, que viene refrendado por el empleo de algunas de las siguientes técnicas:

- *Redes neuronales artificiales:* Son modelos predecibles no lineales que aprenden a través del entrenamiento.
- *Árboles de decisión:* Estructuras arborescentes que representan conjuntos de decisiones que generan reglas para la clasificar un conjunto de datos.
- *Método del vecino más cercano:* Clasifica cada registro basándose en una combinación de los registros de datos históricos más similares a él.
- *Regla de inducción:* Se produce una extracción de reglas del tipo *if-then* de datos basándose en criterios estadísticos.

Estas técnicas no son novedosas y han estado utilizándose durante más de una década en herramientas de análisis especializadas que trabajaban con volúmenes de datos relativamente pequeños. Ahora están evolucionando para inte-

grarse directamente con herramientas OLAP[3]. Aun así, el proceso de minería de datos todavía tiene numerosos aspectos de mejora a considerar, como puedan ser la elaboración de métricas de evaluación para analizar los resultados alcanzados o la importancia de efectuar un análisis considerando datos que varían en tiempo real (Howard, 2013).

1.2. Ciclo de vida de Big Data

Al igual que otros procesos tecnológicos, la ejecución del proceso de Big Data tiene lugar a lo largo de un ciclo de vida donde se selecciona un subconjunto de datos, que no puede modificarse, y se procede a su procesamiento en sucesivas etapas. Si fuese preciso modificar este subconjunto de datos, sería necesario realizar una nueva iteración del proceso. El ciclo de vida a ejecutar se desarrolla en cinco fases o etapas:

1) *Adquisición de datos*: Los datos que se emplean en Big Data provienen de actividades interesantes para la sociedad. Estos datos pueden captarse y almacenarse a través de sensores, simulaciones o transacciones que efectúen los ciudadanos. La frecuente diversidad de estos datos hace necesario que sean filtrados y organizados de forma correcta para su posterior procesamiento.

2) *Extracción y Limpieza de Información*: Normalmente, la información recolectada es una información superior a la que realmente se necesita. Por tal motivo, es preciso efectuar una extracción de la información que permita disponer únicamente de la información necesaria y que ésta sea almacenada de forma estructurada. Además, estos conjuntos de datos suelen contener ausencias de información o detalles inútiles, conllevando a que, muchas veces, el conocimiento encerrado en ellos sea incompleto, lo que obliga a realizar una limpieza del conjunto de datos. Esta limpieza se hace especialmente necesaria si el conjunto de datos proviene de la Web, donde técnicas como *Linked Data* permiten estructurar los datos de forma correcta para su posterior procesamiento. (Wood *et al*., 2014). Esta fase de extracción y limpieza también se conoce con el nombre de reducción de datos, y es un proceso complejo que debe tener la habilidad de no eliminar información que pudiera ser relevante para Big Data.

3) *Integración y Conversión de Datos*: Con frecuencia, la información recolectada proviene de múltiples fuentes heterogéneas que tienen su propio formato de almacenamiento, provocando que el formato de la muestra sea inadecuado al necesario para afrontar el análisis. Para conseguir integrar y convertir los datos de

3 *On Line Analytical Processing*

la muestra se emplean una serie de técnicas de transformación que estructuran y convierten los datos a un formato útil para los algoritmos a emplear en el análisis. La complejidad y duración de esta fase de conversión e integración depende, sobremanera, de cómo ha sido diseñada la base de datos y si se ha considerado un almacenamiento de datos universal que pueda utilizarse en diferentes aplicaciones.

4) *Análisis y Modelización*: Los métodos que se emplean a la hora de analizar y modelizar el proceso de Big Data difieren de los métodos tradicionales estadísticos que se suelen emplear cuando se dispone de una muestra pequeña. En este caso, la muestra de Big Data se caracteriza por ser dinámica, heterogénea, interrelacionada y con cierta presencia de ruido, pero aún así, es una muestra cuyo análisis proporciona información de gran valor, al trabajar con un mayor volumen de información y obtener patrones y conocimiento hasta ahora desconocidos.

5) *Interpretación*: Esta fase recibe los resultados obtenidos en el análisis y requiere de una interpretación por parte de la persona que tiene que tomar las decisiones. En esta interpretación se descartan las reglas carentes de valor y se conservan las tendencias o asociaciones que puedan ser de mayor utilidad en su toma de decisiones. Actualmente, esta etapa interpretativa está siendo ejecutada por un grupo de personas, quienes debaten la importancia de los resultados alcanzados para evitar que se descarte algún resultado relevante.

1.3. Privacidad de los datos

A pesar de los beneficios que aporta esta nueva forma de razonamiento deductivo, no está exenta de problemas. La automatización de los procesos de carga de datos, la integración de datos provenientes de fuentes heterogéneas o la presencia de ruido son algunos de los problemas que se plantean dentro del ámbito tecnológico.

Sin embargo, el problema que más preocupa a la sociedad es el impacto que pueden tener en la privacidad de los ciudadanos tecnologías como Big Data. Esta tecnología supone una fuente de innovación que genera grandes beneficios sociales, pero también plantea múltiples retos de privacidad y seguridad que no pueden ignorarse.

Desde hace unos años la web social y los dispositivos móviles propician la circulación de un inmenso caudal de datos personales en todas direcciones: fotos etiquetadas, tweets, vídeos virales, whatsapps o conexiones a wifis públicas son alguna de las interacciones que los ciudadanos realizan diariamente con toda normalidad sin tener en cuenta las repercusiones sobre su seguridad y privacidad (Borglund; Engvall, 2014).

Cerca del 84% de los usuarios manifiestan su rechazo a proporcionar información personal si desconocen para qué va a utilizarse. Sin embargo, el uso y explotación de estos datos es algo habitual en plataformas y redes sociales como Facebook, donde dichos datos se emplean para establecer criterios de segmentación que puedan utilizar sus colaboradores. Esta creación de perfiles de consumidores puede llegar a convertirse en anhelo para los hackers presentes en la web de cara a una futura comercialización.

Big Data, al trabajar con perfiles de personas extraídos de los datos, permite obtener conclusiones sobre ciudadanos o predecir la probabilidad de encontrarse en determinadas situaciones, que pueden generar una enorme repercusión sobre su privacidad. Es necesario establecer una salvaguarda que garantice el uso que se va a efectuar de los datos proporcionados, evitando que los datos acaben siendo destinados a usos muy diferentes de los originalmente previstos traspasando los límites de lo ético y de lo legal.

Los ciudadanos quieren creer que el desarrollo de aplicaciones tecnológicas se efectúa considerando unas ciertas normas de privacidad. Sin embargo, si se comparte información personal se pierde la privacidad y la privacidad implica, también, un control de la información personal y como se utiliza. Por tanto, es preciso resaltar que la tecnología no va a reemplazar la necesidad de leyes o normas sociales que protejan la privacidad de las personas.

En este sentido el *Massachusetts Institute Technology* indica que, a día de hoy, para alcanzar el objetivo de desarrollar aplicaciones que garanticen la privacidad, es necesario focalizar más en el desarrollo de pautas o mecanismos que protejan los derechos del consumidor y regularicen el uso de los datos por parte de las aplicaciones que emplean tecnología Big Data (MIT, 2014).

2. APLICACIONES EN LA SOCIEDAD

Como bien ha recalcado en alguna ocasión el profesor Arranz:

> Si algún acontecimiento ha marcado el progreso de las sociedades avanzadas, sobre todo en las últimas décadas, ha sido el imparable desarrollo y uso masivo de las herramientas tecnológicas para realizar todo tipo de tareas (Arranz *et al*., 2004).

Aunque la genómica o la astronomía han sido los primeros campos que han trabajado con algoritmos de Big Data, es preciso observar la importancia que esta tecnología está alcanzando en sectores de referencia para la sociedad, como son los servicios, la educación, o el sector gubernamental. Por tanto, se va a detallar, de forma más concreta, los progresos y esperanzas que se tiene depositadas en esta tecnología con respecto a dichos sectores.

2.1. Sector servicios

Las primeras implantaciones comerciales de la tecnología Big Data han tenido lugar dentro del sector servicios. Bancos, empresas de telecomunicaciones, grandes cadenas de alimentación, compañías de transportes o cadenas hoteleras han sido los primeros que comenzaron a ser conscientes del gran tesoro, a modo de información, que residía en sus almacenes de datos.

Las entidades bancarias descubrieron que tomando las transacciones de sus clientes les permite clasificarlos en categorías o establecer perfiles de comportamiento, para efectuar una toma de decisiones basada en datos y no en la intuición humana. La búsqueda de patrones o secuencias de actividades que ayuden a detectar posibles fraudes o la morosidad de sus clientes hacen que esta tecnología sea la principal asesora en la toma de decisiones.

Por su parte, las empresas de telecomunicaciones o las grandes cadenas de alimentación o transportes utilizan la tecnología Big Data para conocer los hábitos de compras de sus clientes y emplear esta información para incrementar sus ventas o mejorar la calidad de sus servicios. Las transacciones de los clientes se emplean en técnicas de marketing para modelizar el comportamiento del consumidor y construir un modelo predictivo basado en datos históricos de consumo que permite clasificar a los potenciales clientes y así conocer y predecir su secuencia de actividades. (Baldi *et al*., 2003)

Las expectativas del sector están depositadas, actualmente, en el conocimiento que puedan llegar a extrapolar del cliente a partir de los repositorios de datos provenientes del comercio electrónico.

2.2. Área educativa

Una prueba de la importancia que está alcanzando el fenómeno de Big Data es su aplicación por parte de las instituciones educativas quienes están empezando a explotar y comprender las ventajas que les ofrece. Si hasta ahora, eran las empresas y organizaciones quienes llevan años analizando estos enormes conjuntos de datos para conocer mejor a sus clientes y predecir las tendencias del mercado, "el mundo educativo está empezando a integrar los conjuntos de datos de que dispone para mejorar el proceso de aprendizaje de los alumnos" (Sanchez, 2014).

Estas nuevas iniciativas han conducido a que numerosas instituciones educativas analicen los datos provenientes de las interacciones de sus alumnos. Este análisis proporciona una serie de conclusiones que van a permitir mejorar el

entorno de trabajo, generando nuevas maneras de estructurar las organizaciones educativas, y lo que es más importante, nuevas formas de aprender.

En este mismo sentido, el NMC Horizon Report, un referente a nivel mundial de las tendencias tecnológicas emergentes en educación, prevé en su último informe de 2014 que "el análisis de datos se adoptará de manera significativa en un plazo de entre dos y tres años, y que de hecho ya se está utilizando en algunas universidades americanas como la de Connecticut" (Johnson *et al*., 2014).

Esta adopción vendrá apoyada por la rápida implantación de los entornos virtuales de aprendizaje y de los MOOC (*Massive Open Online Course*), donde los estudiantes realizan las tareas online dejando un rastro de datos en la web. La recolección y análisis de estos datos provenientes de las transacciones que efectúan los alumnos al interactuar con el sistema se utilizará para adaptar los contenidos a las necesidades y características de los alumnos y así actuar en la mejora del sistema educativo.

Esta mejora del sistema educativo se está realizando basándose en los resultados que proporcionan los sistemas analíticos, conocidos más concretamente como *learning analytics*. Los modelos basados en estos sistemas analíticos están orientados, en su diseño y concepción, a informar y potenciar tanto las actividades de los docentes como de los estudiantes (Baker, 2009).

Los docentes deben observar los patrones de comportamiento generados y así, reducir el riesgo de abandono de los alumnos, a través de un proceso de aprendizaje más personalizado. Igualmente, las analíticas educativas permitirán detectar nuevos problemas, generando posibles correcciones que mejoren el proceso de enseñanza-aprendizaje, o, incluso, cuestionarse la eficacia de los programas docentes que se imparten en el centro educativo.

De igual forma, los estudiantes también salen beneficiados, ya que, gracias al análisis de estos datos, los profesores pueden adaptar los entornos de aprendizaje a sus necesidades. Esta adaptación del entorno dependerá de la creatividad del docente, quien interpretará los patrones de cada alumno y deberá optar por aportar soluciones creativas que ayuden al estudiante a aprender los conocimientos necesarios de la materia. (Trevitt et al., 2012)

Dentro del área educativa, las principales técnicas que se están utilizando para efectuar la toma de decisiones son:

- *Predicción*: Desarrolla un modelo de inferencia a partir de los datos. Se emplea para simular el comportamiento de los estudiantes en función de las actividades previas realizadas y para predecir posibles inconsistencias futuras.
- *Agrupamiento*: Busca clasificar datos o registros en grupos que presenten las mismas características. Permite establecer patrones comunes entre estudiantes que estén englobados en el mismo grupo.

- *Asociaciones de datos*: Procesan los datos para descubrir asociaciones ocultas entre los datos. Permite establecer una asociación de actividades que permiten inducir un secuenciamiento de actividades.
- *Visualización*: Muestra tendencias en el uso de las plataformas educativas que detectan las desviaciones respecto al proceder medio de la clase.

La importancia del impacto que Big Data está teniendo en el sector educativo está empezando a verse reflejada en la expectación suscitada en un gran porcentaje de profesores e investigadores que tienen puestas sus esperanzas en que el análisis proporcione datos relevantes y lo que el uso de estos datos pueda significar para el ámbito de la educación. El empleo de técnicas de Data mining junto con los modelos utilizados en *learning analytics* pueden inducir nuevos sistemas de aprendizaje online y resaltar aspectos educativos desconocidos sobre los que no se estaban haciendo ningún tipo de actuación, a pesar de ser significativos para el proceso de enseñanza-aprendizaje (Ferguson, 2012).

Una muestra de la aplicación de las *learning analytics* en el ámbito educativo puede observarse en la investigación de la profesora Olga Arranz quien, trabajando en la Universidad Pontificia de Salamanca con una muestra superior al millón de registros, ha analizado el uso que los estudiantes universitarios hacen de las herramientas disponibles en los entornos virtuales de aprendizaje (Arranz; Alonso, 2013).

2.3. La administración y Big Data

Big Data está representando una gran oportunidad para los gobiernos actuales a la hora de compartir con los ciudadanos el conocimiento acumulado en sus bases de información (Jagadish *et al*., 2014). Esta aplicación de Big Data, conocida como Open Data, puede definirse como:

> Una filosofía y práctica que persigue que determinados datos e informaciones pertenecientes a las Administraciones Públicas sean accesibles y estén disponibles para todo el mundo, sin restricciones técnicas ni legales, en formatos digitales, estandarizados y abiertos, siguiendo una estructura clara que permita su manejo y comprensión (Piedrabuena; Criado, 2012).

De esta forma, se oferta una nueva fuente de información fiable e integra, para su uso por parte de aplicaciones que accedan desde la Cloud, los dispositivos móviles o los portales sociales (O'Brien, 2012) Para que esta oferta sea real es preciso que se cumplan algunas pautas desde el punto de visto tecnológico:

- Pasar los registros administrativos en formato papel a formato electrónico para mejorar la compartición y accesibilidad de la información.
- Digitalizar los archivos, imágenes y mensajes de manera que permitan efectuar un control y una gestión conforme a las normas estipuladas.
- Establecer un formato universal de los datos para evitar perdidas de tiempo en costosos procesos de cambio de formatos.

Uno de los sectores bajo control gubernamental donde la tecnología Big Data está despertando más expectativas es el sector sanitario. Estas expectativas están fundamentadas en el alto potencial de desarrollo que se puede alcanzar con el análisis de los datos procedentes de los miles de expedientes acumulados en los hospitales. El procesamiento de estos datos, clasificados por enfermedades, permitirá orientar las investigaciones hacia factores hasta ahora desconocidos en busca de una alternativa mejor a las aplicadas que permitan avanzar en el tratamiento de las enfermedades (Ruther, 2014). Avances en el tratamiento del cáncer o la búsqueda de asociaciones entre enfermedades o síntomas son algunas de las principales áreas de aplicación.

Si la implantación en el análisis de enfermedades está en auge, en el campo de la biotecnología se están consiguiendo resultados realmente espectaculares, donde el secuenciamiento de la cadena de ADN está permitiendo la elaboración de profundas investigaciones que repercutan en un descubrimiento mayor de la biología del ser humano.

A pesar de todo, el impacto de Big Data en los departamentos gubernamentales está dando los primeros pasos y es todavía relativamente pequeño para el alto potencial que tiene. Sólo cuando los gobernantes, escépticos aún acerca de sus ventajas, tomen conciencia de sus posibilidades, se observará un mayor empleo de esta tecnología en beneficio del ciudadano (Shueh, 2014).

CONCLUSIONES Y TENDENCIAS

Como se ha podido ver el futuro es presente. Las nuevas formas de procesamiento de datos están ya implantadas en la sociedad. La importancia que está tomando la tecnología Big Data está alcanzado tales extremos, que ya se está hablando del comienzo de una nueva era. Su alto grado de introducción en sectores relevantes de la sociedad como servicios bancarios, educación, telecomunicaciones, sanidad o en la propia interrelación con los departamentos gubernamentales está empezando a ser tan alta que no es posible quedarse al margen de las futuras implicaciones.

La combinación de esta tecnología con las técnicas de minería de datos aporta un potencial tan grande que es difícil predecir hasta donde puede llegar. Las sugerencias de cambios en las arquitecturas actuales están orientadas a aumentar el potencial de procesamiento y a permitir trabajar con volúmenes incluso mayores de información. La incorporación de la Cloud al proceso de deducción de conocimiento hará que sea necesario implantar un sistema de garantía de calidad de los datos procesados.

La organización inteligente de los datos, el diseño estructurado de los repositorios de los datos, así como la automatización de procesos tecnológicos de carga de datos contribuirán a eliminar el ruido presente y a evitar que se alcancen resultados con un cierto grado de incertidumbre.

De igual forma, es preciso mejorar el tratamiento que se hace de los datos privados de los ciudadanos. Su preocupación por el uso de los múltiples datos que exponen en las redes sociales o en sus transacciones comerciales cuando emplean del ordenador aumenta de tal forma, que se hace necesario establecer normas o pautas de procedimiento que eviten futuras reticencias a la hora de facilitar la información.

Concluyo esperando que esta contribución haya sido del agrado del profesor/filósofo Arranz, y como él diría, en el futuro, *Carpe Diem*.

REFERENCIAS BIBLIOGRAFICAS

ARRANZ, M.; ALONSO, V.; LOPEZ, A. J. (2004), "Formación del Profesorado y TIC. Nuevas Tendencias". En: *Ponencias, Comunicaciones y Talleres del III Congreso Regional de Tecnologías de la Información y de la Comunicación*. Salamanca: Junta de Castilla y León, 45-54.

ARRANZ, O.; ALONSO, V., (2013), "Big Data & Learning Analytics: A Potential Way to Optimize Elearning Technological Tools". En: *Proceedings IADIS International Conference eLearning 2013*. Praga: República Checa: IADIS Press, 313-317.

BAKER, R. (2009), *Data Mining for Education*. Pittsburgh: Carnegie Mellon University, 4-6.

BALDI, P.; FRASCONI, P.; SMYTH, P. (2003), *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*. West Sussex, England: John Wiley & Sons, 212-232.

BORGLUND, A.; ENGVALL, T. (2014), "Open Data? Data, Information, Document or Record?". *Records Management Journal*, Vol. 24, Issue. 2, 2014, 163-180.

DEAN, J. (2014), *Big Data, Data Mining and Machine Learning*. Hoboken, New Jersey: John Wiley & Sons, 3-21.

DIEBOLD, F.X. (2003), "'Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting" (Discussion of Reichlin and Watson papers). En: DEWATRIPONT, M., HANSEN, L.P. and TURNOVSKY, S. (Eds.), *Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society*. Cambridge: Cambridge University Press.

FERGUSON, R. (2012), *The State of Learning Analytics in 2012: A Review and Future Challenges. Technical Report KMI-12-01*. [En línea]. UK: Knowledge Media Institute, The Open University. <http://kmi.open.ac.uk/publications/techreport/kmi-12-01>. [Consulta: 14 mar. 2015].

GIUDICI, P. (2003), *Applied Data Mining: Statistical Methods for Business and Industry*. West Sussex, England: John Wiley & Sons, 209-226.

HERNÁNDEZ, J.; RAMIREZ, M.J.; FERRI, C. (2004), *Introducción a la Minería de Datos*. Madrid: Pearson Educación, 22-25.

HOWARD, H. (2013), *Knowledge Discovery in Databases*. [En línea]. <http://www2.cs.uregina.ca/~dbd/cs831/index.html>. [Consulta: 17 abr. 2015].

IBM, (2013), *Analytics: The Real World Use of Big Data. How Innovative Enterprises Extract Value from Uncertain Data. Executive Report*. [En línea]. IBM Institute for Business Value. < http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=XB&htmlfid=GBE03521USEN#loaded >. [Consulta: 17 abr. 2015].

JAGADISH, H.; GEHRKE, J.; LABRINIDIS, A.; PAPAKONSTANTINOU, Y.; PATEL, J.; RAMAKRISHNAN, R.; SHAHABI, C. (2014), "Big Data and its Technical Challenges". *Communications of the ACM*, Vol. 57, Num. 7, 2014, 86-94.

JOHNSON, L.; ADAMS, S.; ESTRADA, V.; FREEMAN, A. (2014), *NMC Horizon Report: 2014 K-12 Edition*. Austin, Texas: The New Media Consortium, 40-42.

JOYANES, L. (2014), *Big Data: Análisis de Grandes Volúmenes de Datos en Organizaciones*. México: Mancorbo, 327-344.

MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; HUNG, A. (2011), *Big Data: The Next Frontier for Innovation, Competition, and Productivity. Executive Summary.* [En línea]. McKinsey Global Institute. <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>. [Consulta: 14 abr. 2015].

MIT, (2014), *Big Data Privacy Workshop: Advanced the State of the Art in Technology and Practice.* [En línea]. Cambridge: Massachusetts Institute Technology,. <http://web.mit.edu/bigdata-priv/images/MITBigDataPrivacyWorkshop2014_final05142014.pdf>. [Consulta: 28 abr. 2015].

O´BRIEN, A. (2012), *The Impact of Big Data on Government*. [En línea]. IDC Government. <http://www.ironmountain.com/Knowledge-Center/Reference-Library/View-by-Document-Type/White-Papers-Briefs/Sponsored/IDC/The-Impact-of-Big-Data-on-Government.aspx>. [Consulta: 26 abr. 2015].

PIEDRABUENA, A.; CRIADO, L. (2012), "OpenData, Oportunidad Escondida y Semilla de la Web Semántica". *RUIDERAe: Revista de Unidades de Información*, Num. 2, 2012, 1-21.

RUTTER, T. (2014), *How Big Data is Transforming Public Services*. [En línea]. <ww.theguardian.com/public-leaders-network/2014/apr/17/big-data-government-public-services-expert-views>. [Consulta: 26 abr. 2015].

SÁNCHEZ, O. (2014), "Learning Analytics, el Big Data en Versión Educativa". En: *El Blog d'UPCnet*. [En línea]. <http://blog.upcnet.es/learning-analytics-el-big-data-en-version-educativa/>. [Consulta: 6 abr. 2015].

SCHÖNBERGER, V.; CUKIER, K. (2014), *Big Data: A Revolution that will Transform How we Live, Work and Think*. Boston, New York: Mariner Books, 6-12.

SHUEH, J. (2014), *Big Data Could Bring Governments Big Benefits*. [En línea]. <http://www.govtech.com/data/Big-Data-Could-Bring-Governments-Big-Benefits.html#.VTVHnuaPjfg.mailto>. [Consulta: 26 abr. 2015].

TOFFLER, A. (1980), *The Third Wave*. New York, William Morrow and Company.

TREVITT, C.; BREMAN, E.; STOCKS, C. (2012), "Assessment and Learning: Is it Time to Rethink Student´ Activities and Academic Roles?". *Revista de Investigación Educativa*, Vol. 2, Num 30, 2012, 253-267.

WEST, D. (2012), *Big Data for Education: Data Mining, Data Analytics, and Web Dashboards*. [En línea]. Governance Studies, the Brookings Institution,. <http://www.brookings.edu/research/papers/2012/09/04-education-technology-west>. [Consulta: 23 abr. 2015].

WOOD, D.; ZAIDMAN, M.; RUTH, L. (2014), *Linked Data: Structured Data on the Web*. Shelter Island, New York: Manning Publications, 212-232.