

Cómo las tecnologías de análisis de datos pueden ayudar a desarrollar el propósito de la educación de la ciencia y la ingeniería con *big data* en el siglo XXI

Ramón Alfonso Perdomo Salcedo¹
George Victorovich Buslaev²

Resumen: IBM estima que cada día se crean o replican 2,5 quintillones de bytes de datos, esto equivale a un millón de discos duros que se llenan de datos cada hora. En 2015, los centros de datos ocupaban el terreno equivalente a casi 6.000 campos de fútbol. Para 2020 se espera que la cantidad de información digital aumente exponencialmente más de siete veces el volumen existente en 2014. (Desjardins, 2015).

Consideramos, para propósitos prácticos de discusión en este artículo, que la ciencia es un medio para desarrollar explicaciones sobre cómo funciona el mundo natural, y que la ingeniería es un medio para desarrollar soluciones a los problemas humanos. De este modo, diremos que ambas (la ciencia y la ingeniería) están destinadas a mejorar nuestras vidas, lo que representa un fuerte motivador para el desarrollo de un nuevo proceso de aprendizaje en el que podamos dar sentido al análisis de la información (ciencia de análisis de datos) y a la construcción de modelos de validación (análisis descriptivos, predictivos y prescriptivos). Intentaremos desarrollar la siguiente pregunta en el presente trabajo: ¿cómo pueden las técnicas de aprendizaje de análisis de datos ser una fuente de mejora en la calidad de la educación para enfrentar los desafíos de la innovación, la competencia y la productividad en el siglo XXI?

Palabras clave: *big data*, inteligencia artificial, aprendizaje automático, minería de datos, ingeniería, ciencia, educación.

¹ Ingeniero de Petróleos, Estudiante PhD. Tecnologías de perforación y completamiento de pozos. Instituto de Geología, Producción de Petróleo y Gas y Transporte por Oleoducto. Universidad Técnica Estatal de Ukhtá, Ukhtá, Federación Rusa. Correo electrónico: raperdomo@mail.ru

² Ingeniero Mecánico, Profesor Asociado, PhD. Departamento de Perforación. Facultad de Petróleo y Gas. Universidad de minas de San Petersburgo, San Petersburgo, Federación Rusa. Correo electrónico: Buslaev_GV@pers.spmi.ru

Forma de citar: Perdomo Salcedo, R.A. y Victorovich Buslaev, G. (2020). Cómo las tecnologías de análisis de datos pueden ayudar a desarrollar el propósito de la educación de la ciencia y la ingeniería con *big data* en el siglo XXI. *Revista Docencia Universitaria*, 21(1), 19-39.

Recepción: enero 27 de 2020 – **Aceptación:** marzo 15 de 2020

How data analysis technologies can help to develop the purpose of *big data* science and engineering education in the 21st century

Ramón Alfonso Perdomo Salcedo¹

George Victorovich Buslaev²

Abstract: IBM estimates that 2.5 quintillion bytes of data are created or replicated every day. This is the equivalent of one million hard drives filling up with data every hour. In 2015, data centers occupied the equivalent of nearly 6,000 football fields. By 2020, the amount of digital information is expected to increase exponentially by more than 7 times the volume in 2014. (Desjardins, 2015)

We consider, for practical purposes of discussion in this article, that science is a means of developing explanations of how the natural world works, and engineering is a means of developing solutions to human problems. So we will say that both (science and engineering) are aimed at improving our lives, which represents a strong motivator for the development of a new learning process in which we can give meaning to the analysis of information (data analytics) and the construction of validation models (descriptive, predictive and prescriptive analysis). We will try to develop the following question in the present work: How can analytical data learning techniques be a source of improvement in the quality of education to face the challenges of innovation, competition and productivity in the 21st century?

Keywords: *big data*, artificial intelligence, machine learning, data mining, informatics, engineering, science, education.

¹ Ingeniero de Petróleos, Estudiante PhD. Tecnologías de perforación y completamiento de pozos. Instituto de Geología, Producción de Petróleo y Gas y Transporte por Oleoducto. Universidad Técnica Estatal de Ukhá, Ukhá, Federación Rusa. Correo electrónico: raperdomo@mail.ru

² Ingeniero Mecánico, Profesor Asociado, PhD. Departamento de Perforación. Facultad de Petróleo y Gas. Universidad de minas de San Petersburgo, San Petersburgo, Federación Rusa. Correo electrónico: Buslaev_GV@pers.spmi.ru

Introducción. Un mundo basado en datos está cambiando nuestra forma de pensar

Las tecnologías emergentes en el campo de *big data*, minería de datos (*Data Mining*), aprendizaje automático (*Machine Learning*) y la inteligencia artificial (*Artificial Intelligence*), están alimentando la necesidad de planes de estudio innovadores en formatos de entrega en línea, tradicionales e híbridos que satisfagan las necesidades del mercado laboral. Las investigaciones sobre la psicología del aprendizaje y la motivación han descubierto que muchas de las prácticas que componen el contexto de aprendizaje moderno son ineficientes, ineficaces o a veces directamente contrarias a los objetivos de un sistema educativo en evolución. Esto sugiere que estos sistemas, a pesar de su largo linaje, deberían ser reemplazados por aquellos que utilizan una comprensión moderna más completa del aprendizaje y la eficiencia organizacional.

El método educativo tradicional seguido por los programas académicos actuales en los niveles secundario y universitario puede enriquecerse si se incluye el recaudo de información y los análisis se llevan a cabo soportando la matemática y el pensamiento computacional, utilizando argumentos a partir de la evidencia. Las discusiones actuales del uso de herramientas computacionales buscan hacer aportes de investigación pedagógica, materiales y prácticas de enseñanza y mencionar

casos didácticos que aborden la adquisición, aplicación y desarrollo continuo de los conocimientos y habilidades requeridas en el uso del análisis de datos en el aula de clases con énfasis en el análisis de alto rendimiento, la ciencia de datos, la visualización y otras tecnologías analíticas emergentes. A continuación, enumeraremos algunos puntos descritos por [Camins \(2015\)](#) que permiten sintetizar la lógica del proceso de incorporación de análisis de datos (pensamiento analítico) en el aula de clases para los fines de este documento:

1. Hacer preguntas sobre el fenómeno (por ejemplo, las causas de la baja productividad, las causas del aumento de los costos de mantenimiento) y definir los problemas que deben resolverse (diseño de los flujos de trabajo operativos, mejora de la eficiencia operativa).
2. Desarrollar y utilizar modelos. Los modelos representan características relevantes comprobables de las explicaciones científicas o de las soluciones de diseño.
3. Planificar y llevar a cabo investigaciones. Los objetivos de las investigaciones son probar, refinar o reemplazar explicaciones existentes o hipotéticas o soluciones de diseño.
4. Analizar e interpretar los datos.
5. Usar las matemáticas y el pensamiento computacional.

6. Construir explicaciones y diseñar soluciones.
7. Involucrarse en la forma de presentación de argumentos a partir de las pruebas.

Con la explosión de grandes cantidades de datos recaudados de múltiples fuentes, la demanda por aplicaciones de inteligencia artificial y el desarrollo de técnicas de análisis de datos ha aumentado a un mayor ritmo con el paso de los años.

Existe un consenso generalizado en torno a que el valor medible de los datos se crea solo después de su interpretación e implementación en los procesos de la vida cotidiana. Afortunadamente, la educación está cambiando; la revolución de la información, desde la llegada del microprocesador hasta nuestros días, ha cambiado la forma en que pensamos sobre la evaluación y la interpretación de grandes volúmenes de datos. Por ejemplo, los métodos estadísticos que antes eran el dominio de la investigación científica, hoy por hoy se pueden aprovechar mucho mejor en el aula de clases, y la introducción de internet ha proporcionado una plataforma totalmente nueva para aprender y flexibilizar los modernos conceptos de educación virtual.

En un estudio publicado por [Bughin, J. Manyika, J. \(2016\)](#) se comentó que la adaptación a una era en la que la toma de decisiones se centra mayoritariamente en los datos no siempre ha resultado ser una

propuesta sencilla para las personas, las organizaciones y los gobiernos. Muchos sectores están luchando por desarrollar el talento humano necesario para hacer frente a los nuevos retos, los procesos de negocio y el músculo organizacional para capturar el valor real del análisis de datos. Esto se está convirtiendo en un asunto urgente, ya que la capacidad de extraer conocimiento de grandes volúmenes de datos será cada vez más el factor diferenciador en la toma de decisiones soportadas de las empresas, sectores y gobiernos. El liderazgo moderno se orienta a aprovechar y sacar ventaja del volumen, variedad y velocidad con la cual los patrones de información actuales y futuros son recaudados.

Las nuevas tecnologías permiten a las escuelas, colegios, centros de formación y universidades analizar absolutamente todo lo que sucede en el entorno; desde el comportamiento de los estudiantes, los resultados de los exámenes, el desarrollo profesional de los estudiantes, hasta las necesidades educativas de un determinado segmento de población.

Los grandes volúmenes de datos permiten cambios muy interesantes en el campo formativo/educativo que revolucionarán la forma en que los estudiantes aprenden y los profesores enseñan. En este sentido, es importante destacar que la toma de decisiones basada en datos es un componente esencial de los sistemas emergentes de ciencia e ingeniería que generan y

consumen grandes cantidades de datos de detección autónomos.

Es por ello que se dice que los currículos académicos y los programas de estudios deberán centrarse en el uso de tecnologías y metodologías necesarias para inferir información útil e identificar patrones subyacentes a partir de datos a menudo crudos, incompletos, ruidosos y corruptos que están presentes en las aplicaciones de ciencia e ingeniería de la vida real. Las nuevas directrices académicas que incluyan el análisis de datos en las universidades deben ofrecer la oportunidad de explorar soluciones avanzadas de análisis de datos, como la codificación dispersa, la detección compresiva, el filtrado no local y los modelos generativos y discriminatorios de datos.

Big data. Representa la intersección de la estrategia del negocio y la ciencia de datos, ofreciendo nuevas oportunidades para crear ventajas competitivas, permitiendo a los gobiernos, las empresas y a quienes los analicen, la posibilidad de utilizar los datos como un activo estratégico, dotándolos de la información pertinente en tiempo real a la hora de tomar decisiones para eliminar procesos operativos ineficientes, mejorar la experiencia del cliente, aprovechar nuevos mercados, etc.

Big data representa el aprendizaje continuo. Tradicionalmente, las universidades y los centros de formación han basado sus modelos académicos en

el conocimiento en lugar del aprendizaje. Aprender y conocer son dos cosas completamente diferentes: el aprendizaje viene con la experiencia y está asociado con la asunción de riesgos. Saber es más bien un conocimiento previo (es decir, el “instinto”) o, tal vez, el resultado de haber sido “dicho” un cierto hecho es cierto sin cuestionarlo en la cultura del conocimiento. Conocimiento es el inverso del aprendizaje, a menudo descubrimos los hechos tras realizar los análisis pertinentes. Es necesario que las universidades y centros de formación abran el camino a un nuevo modelo educativo en el que la cultura del conocimiento sea complementada por la cultura del aprendizaje.

Big data: una innovación disruptiva

¿Por qué *big data* es diferente de cualquier otra moda que hayamos tratado en el pasado? Para responder esta pregunta, debemos entender que la sociedad actual genera datos. Hay “cuatro V’s” que caracterizan estos datos: Volumen, Velocidad, Variedad y Veracidad. [Ravindra, S. \(2017\)](#) confirma que *big data* es una innovación que crea una nueva red de valor y mercado, y que puede alterar por completo un mercado existente. Entonces respondemos a la pregunta afirmando que una red de valor al desplazar a las alianzas, productos y empresas líderes y altamente establecidas se conoce como una innovación disruptiva. [Christensen \(1997\)](#) y sus colaboradores definieron y analizaron el fenómeno de innovaciones

disruptivas el año en su libro “El dilema del innovador: cuando las nuevas tecnologías hacen fracasar a las grandes empresas”. En esencia, debemos aclarar que toda innovación revolucionaria no es disruptiva. Cuando una revolución crea un desorden en el mercado y sus agentes, entonces se considera sin ninguna duda como disruptiva.

El término “innovación disruptiva” ha sido muy popular en los últimos años cuando hablamos de tecnologías y emprendimientos, a pesar de las muchas diferencias en la aplicación, muchos están de acuerdo en lo siguiente.

Las innovaciones disruptivas son:

- Más accesibles (con respecto a la distribución o usabilidad).
- Más baratas (desde la perspectiva del cliente).
- Un modelo de negocio con ventajas de costes estructurales (con respecto a las soluciones existentes).

En el aula de clases, los grandes volúmenes de datos son una fuerza disruptiva y nos recuerdan que las personas necesitan algo más que nuevas habilidades, tecnologías y herramientas. Las personas que se enfrentan a *big data* en los centros educativos necesitan una mente abierta para repensar los procesos educativos que han seguido durante mucho tiempo y transformar la forma en que aprenden. Sin embargo, no es particularmente fácil forzar este tipo de

cambio hacia el pensamiento analítico en los estudiantes tradicionales. El objetivo general de *big data* dentro del sistema educativo debería ser mejorar los resultados de los estudiantes. Los mejores estudiantes son buenos para la sociedad, las organizaciones y las instituciones educativas. [Rijmenam \(2013\)](#) afirma que “el uso de *big data* en entornos de clase genera cambios significativos en el rendimiento de los estudiantes y en la flexibilidad de las metodologías de aprendizaje”. La innovación disruptiva que proviene de los grandes datos son los grandes procesos y tecnologías de análisis de datos:

- Al mejorar los resultados de los estudiantes y al potenciar su comprensión.
- Al crear programas personalizados en masa.
- Al mejorar la experiencia de aprendizaje en tiempo real.
- Al reducir las deserciones y aumentar la motivación y el compromiso de los estudiantes con los resultados.

El volumen actual de datos creados es sustancial: tanto es así que el 90% de los datos mundiales han sido creados en los últimos dos años, por lo que la revista *The Economist* estimó, como ya mencionamos anteriormente, que habrá aproximadamente siete veces más datos en 2020 que en 2014 ([Desjardins, 2015](#)). Estos conjuntos de datos extremadamente grandes pueden

ser analizados computacionalmente para revelar patrones, tendencias y asociaciones, especialmente en relación con los comportamientos e interacciones humanas.

Análítica del conocimiento, la visión tradicional del mundo académico en ciencia e ingeniería

En el mundo académico actual las prioridades estratégicas del sistema educativo en su conjunto tienden a centrarse en cumplir el marco regulatorio y seguir el programa de estudios y créditos aprobado. Como resultado, la analítica de datos que se enseña en las universidades a menudo se centra demasiado en matemáticas y estadísticas complejas en lugar de la enseñanza de la creación de valor a largo plazo. Hemos incorporado el concepto “Desarrollo Sostenible” a nuestro lenguaje cotidiano, pero cuantos de nosotros comprendemos que la creación de sostenibilidad viene dada por la búsqueda de valores de largo plazo que están presentes en entornos cambiantes y algunas veces adversos. Sostenibilidad es pues, flexibilidad y capacidad de adaptación a realidades cambiantes.

Las capacidades de almacenamiento y análisis de datos han dado enormes saltos en los últimos años. El volumen de datos disponible para análisis ha crecido exponencialmente, con el paso del tiempo se han desarrollado algoritmos más sofisticados y la potencia y el almacenamiento de los sistemas computacionales mejoran constantemente.

La convergencia de estas tendencias está impulsando rápidos avances tecnológicos y disrupciones constantes en todo tipo de modelos organizacionales. Incluso en la forma en que los gobiernos se relacionan con los ciudadanos.

Humby (1994), de Tesco, afirmó: “La información es el petróleo del siglo XXI”, y Sondergaard (2011) lo complementó diciendo “y el análisis es el motor de combustión”. Sondergaard (2011) citado por Pettery (2011) comentó también: “perseguirlo estratégicamente creará una cantidad sin precedentes de información de enorme variedad y complejidad”. Esto está llevando a un cambio en las estrategias de gestión de datos conocido como *big data*.

Si explicamos la definición de “modelo académico” – como el patrón conceptual a través del cual se esquematizan las partes y los elementos de un programa de estudios, lograremos fácilmente comprender que “los modelos académicos” deben variar de acuerdo al periodo histórico en que se contextualizan, ya que su vigencia y utilidad dependen del contexto social.

Por esta razón describiremos brevemente los diferentes criterios de combinación de los modelos académicos en el pasado reciente y en el presente, propuestos por [Minelli et al. \(2013\)](#), para entender sus contextos.

- Negocios + matemáticas dio origen a la profesión de consultor.

Esto nos permitió usar la heurística y la creatividad para hacer argumentos persuasivos en la sala de juntas.

- Negocios + tecnología dio lugar a las tecnologías de la información. Esto nos ayudó a automatizar las tareas algorítmicas, mejorando así la productividad y la eficiencia.
- Matemáticas + tecnología inspiró interesantes productos de software que nos ayudaron a abordar una amplia gama de problemas empresariales y a operar proactivamente con anticipación.

Entonces para describir mejor la orientación de las competencias necesarias para la creación de empresas y modelos académicos en el desafiante contexto actual y futuro cercano podemos hacer referencia al enfoque propuesto por [Minelli et al. \(2013\)](#).

- Negocios + matemáticas + tecnología, junto con las ciencias del comportamiento, nos permiten conectar los puntos entre interacciones y desarrollar una comprensión más profunda del comportamiento humano. Combinando ese profundo entendimiento con las matemáticas y la tecnología podremos crear los incentivos apropiados para impulsar comportamientos que se alinean con las metas de negocio.

Las tres etapas más importantes en la evolución de los sistemas de datos

Big data es la próxima generación de almacenamiento de datos y análisis de negocio. El conocimiento obtenido a partir de grandes volúmenes de bases de datos es la orientación requerida para los actores de mercado que busquen obtener ingresos y garantizar la sostenibilidad de su modelo de negocios de manera eficiente. La mayor parte de este fenómeno está dado por el rápido ritmo de innovación y cambio de las sociedades modernas; donde estamos hoy no es donde estábamos hace cinco años y definitivamente no será donde estaremos en una década.

Ghosh (2008) citado por [Minelli et al. \(2013\)](#) explica: “Aparte de los cambios en la tecnología actual de hardware y software, también ha habido un cambio masivo en la evolución actual de los sistemas de datos” (capítulo 1, página 2), Ghosh lo comparó con las etapas del aprendizaje: “dependiente, independiente e interdependiente”. Estos conllevan a los análisis de datos y la construcción de modelos de validación mostrados en la [figura 1](#).

- Dependiente (primeros días). Los sistemas de datos eran bastante nuevos y los usuarios no sabían muy bien lo que querían. Se colectaba un gran volumen de datos, muchas veces sin propósito específico.
- Independiente (años recientes). Los usuarios entendieron lo que era una plataforma analítica y

trabajaron junto con el desarrollo de las TIC para definir las necesidades del negocio y el enfoque para obtener información de los usuarios, el mercado, el entorno o las operaciones para la empresa.

- Interdependiente (la era del *big data*). Etapa interactiva entre varias empresas, donde se crean y fortalecen los vínculos de colaboración más allá de las paredes de la empresa.

La *figura 1* describe los diferentes tipos de análisis necesarios:

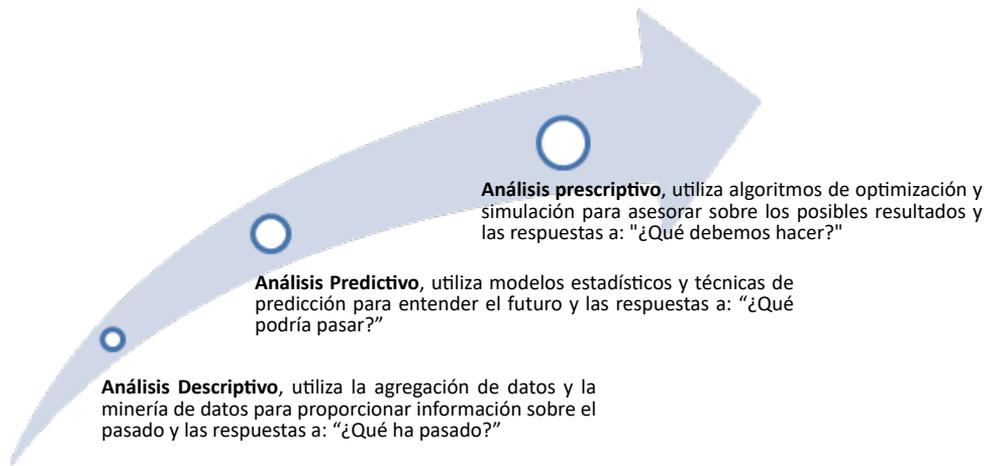


Figura 1. Tipos de análisis.

Fuente: Los autores.

Analítica de entornos para una correcta alineación de los programas académicos con las necesidades del mercado

Cada economía y sector económico tiene sus propias particularidades, que resulta imperativo comprender para diseñar programas académicos que respondan a las necesidades de la sociedad en la que interactúan los profesionales graduados. Es por ello que vamos a analizar rápidamente tres

ejemplos, mostrando las diez empresas más valiosas (por valor de mercado). En los dos primeros ejemplos, analizaremos la estrategia de los dos principales países definidos como potencias económicas del planeta (Estados Unidos y China) para establecer cómo son sus dinámicas internas de mercado orientadas hacia la generación del análisis de datos en los principales sectores productivos. En el tercer caso, tomaremos como ejemplo el caso del sector económico minero energético y mostraremos sus

propias dinámicas hacia la orientación del desarrollo de análisis de datos con aplicaciones.

La metodología vincula el valor del mercado de empresas abiertas (que

cotizan en bolsa y democratizadas) con la concentración de los sectores productivos en los que participan para determinar la orientación, el impulso del big data y el análisis de datos, en la reforma de sus programas y sus modelos académicos.

Tabla 1.

Ranking de las 10 empresas más valiosas por valor de mercado de Estados Unidos en 2018.

Posición	Compañía	Sector	Valor de mercado (USD millardos)
1	Apple	Computadores, oficina, equipos	926,9
2	Amazon	Servicios de Internet y venta al por menor	777,8
3	Alphabet (Google)	Servicios de Internet y venta al por menor	766,4
4	Microsoft	Computadores, software, B2B.	750,6
5	Facebook	Empresa de servicios de redes sociales	541,5
6	Berkshire Hathaway	Seguros: Propiedad y siniestro (acciones)	491,9
7	JP Morgan Cheese	Banco de inversión	387,7
8	Exxon Mobil	Petroleo & Gas	344,1
9	Johnson & Johnson	Farmacéutica	341,3
10	Bank of America	Banco Comercial	313,5

Fuente: (Statista, 2019; IMF,2019)

En *ranking* de la *tabla 1* mostramos las diez empresas de mayor valor de mercado en los Estados Unidos para el 2018 (US \$ 5641,7 Millardos), al revisarla, encontramos una gran concentración de empresas en el negocio de tecnología digital que básicamente se concentran en el comercio electrónico y el conocimiento de los hábitos de compra de sus usuarios. Para explicar mucho mejor el crecimiento del valor de la información en el mercado estadounidense, podemos

decir que las empresas bajo el acrónimo GAFAT (Google, Apple, Facebook, Amazon, Twitter) (USD\$ 3 040.09 millardos) agregaron casi la misma capitalización de mercado que el PIB de un país como Guyana (USD\$ 3 636 millardos). Las empresas de GAFAT y su entorno, el valle del silicio (Silicon Valley), se han convertido en maestros en la explotación del valor creado por el cliente a partir de la combinación definida anteriormente: Negocios + Matemáticas

+ Tecnología, junto con las ciencias del comportamiento.

Entonces, para entender la dinámica del mercado americano y la respuesta dada por los centros académicos y de investigación es necesario entender lo que ha sucedido en las últimas dos décadas; en Estados Unidos han surgido comunidades de conocimiento que se han fortalecido con la participación y el soporte de la academia y el sector productivo, las *start-ups* han surgido como los maestros de la innovación, sus productos y tecnologías innovadoras no solo han cambiado profundamente la vida de las personas, sino que también han ejercido una presión considerable sobre los actores corporativos establecidos. Cuando alcanzan la madurez, la mayoría de las industrias frenan su crecimiento o incluso dejan de crecer. En este ambiente las nuevas empresas amenazan con alterar las viejas industrias con nuevas innovaciones, por lo que las grandes corporaciones más que nunca se han visto obligadas a estrechar sus vínculos con la academia para innovar y crear nuevas líneas de negocio para sobrevivir, así como mejorar sus reconocimientos tempranos de emprendimientos disruptivos para adquirirlos e incorporarlos, especialmente, dado el hecho de que el ritmo de irrupción de las nuevas empresas se está acelerando. Comentamos aquí, como las nuevas empresas de tecnología basada en software como Amazon, Facebook, Google, Tesla y Uber, que eran pequeñas empresas de nueva creación hace tan

solo unos años, están amenazando hoy a los actores establecidos y a sus mercados principales. Los centros universitarios en muchos casos han respondido estimulando los centros de innovación y los convenios con los sectores productivos, flexibilizando los requerimientos de titularización (educación virtual) y ofreciendo planes y programas orientados a capacitar personal de sectores específicos (desarrollando programas, a medida, para sus asociados en los sectores de interés).

En la lista de la *tabla 2* mostramos las diez empresas de mayor valor de mercado en la República Popular China para el 2018 (China \$ 2 638,4), al revisarla, encontramos una gran concentración de empresas en tres sectores: principalmente banca, comercio electrónico y tecnología, y petróleo y gas.

Debemos comentar que China es el país de más rápido crecimiento económico en el mundo desde la década de 1980. China es el centro mundial para la fabricación de todo tipo de productos y la indiscutible mayor potencia industrial y exportadora de bienes a nivel mundial. China es el país más poblado del mundo y con la mayor tasa de crecimiento en consumo, además de segundo mayor importador de mercancías. Por tanto, resulta lógico pensar que los esfuerzos del gobierno Chino se centren en el fortalecimiento de sus transacciones financieras (Banca), en un aumento de

sus exportaciones (Comercio electrónico) y en garantizar la provisión energética para su modelo de desarrollo (Petróleo y Gas).

Es por ello, que hoy en día, la mayoría de las organizaciones bancarias, de servicios financieros y de seguros (BFSI) están trabajando arduamente

con las universidades para adoptar un enfoque totalmente basado en datos para hacer crecer sus negocios y mejorar los servicios que proporcionan a sus clientes (bancarización). Al igual que la mayoría de las otras industrias, el análisis de datos es un factor crítico para los que trabajan en el sector financiero y comercial en China.

Tabla 2.

Las diez empresas más valiosas por valor de mercado de China (2018)

Posición	Compañía	Sector	Valor de mercado (USD millardos)
1	Alibaba	Comercio, venta al por menor, internet	499,4
2	Tencent Holdings	Servicios y productos relacionados con Internet.	491,3
3	ICBC	Banco industrial y comercial	311,0
4	China Construction Bank	Banco comercial	261,2
5	PetroChina	Petróleo & gas	220,2
6	China Mobile	Servicios móviles de voz y multimedia	192,6
7	Agricultural Bank of China	Banco comercial	184,1
8	Ping An Insurance Group	Compañía de seguros	181,4
9	Bank of China	Banco comercial	158,6
10	Sinopec	Petróleo & gas	138,6

Fuente: (Statista, 2019; IMF, 2019)

La estrategia nacional de desarrollo económico, social y productivo, va de la mano con el establecimiento de programas educativos que fomentan la ciencia del análisis de datos en China. El gobierno Chino ha priorizado la

revitalización de zonas rurales, la inteligencia artificial, la seguridad en redes y la enseñanza de idiomas extranjeros como pilares para los futuros programas de pregrado y postgrado en sus universidades.

Beijing planea ser el líder mundial en la tecnología de *big data* para el año 2030. Esto se puede ilustrar en que China es abanderada en los sistemas de reconocimiento facial, utilizados por las fuerzas del orden. China ha perfeccionado el sistema de retratos más grande del mundo que cubre a más de 1500 millones de personas. Este es uno de los múltiples casos donde aplicaciones de *big data* muestran sus características de innovación disruptiva.

Blystone (2019) comentó que “a nivel de comercio, por ejemplo, mucha gente sabe que Alibaba es un

minorista chino en línea”, pero el modelo de negocio de la compañía es menos conocido y sorprendentemente diferente del líder del comercio electrónico en los Estados Unidos, Amazon.com, Inc. Poca gente sabe que en el caso de Alibaba tres grandes portales web constituyen el núcleo del negocio de Alibaba: *business* Alibaba, Taobao y Tmall. Estos tres sitios web de comercio electrónico sirven para conectar varios tipos de compradores y vendedores, en los que Alibaba actúa como un agente intermediario. Esa es pues, la característica del enfoque del entorno chino de los análisis de datos.

Tabla 3.

Diez empresas más valiosas por valor de mercado sector petróleo y gas, minería (2018)

Posición	Compañía	Sector	Valor de mercado (USD millardos)
1	ExxonMobil	Petroleo & gas	344,1
2	Royal Dutch Shell	Petroleo & gas	306,5
3	Chevron	Petroleo & gas	248,1
4	PetroChina	Petroleo & gas	220,2
5	Total	Petroleo & gas	168,0
6	BP	Petroleo & gas	152,6
7	Sinopec	Petroleo & gas	138,6
8	BHP Billiton	Minería, metales, petroleo & gas	133,2
9	Rio Tinto	Minería	99,3
10	Schlumberger	Servicios petroleros	98,5

Fuente: (Statista, 2019; IMF, 2019).

En la lista de la *tabla 3* mostramos las diez empresas de mayor valor de mercado en el sector minero energético para el 2018 (Petróleo y gas, minería USD\$1909,1 millardos). Realizaremos

este breve estudio para comprender de qué manera las empresas del sector extractivo (Petróleo y gas, minería) han venido desarrollando sus conceptos propios de arquitectura de valor. Cuando hablamos

de arquitectura de valor nos basamos en la definición de [Holland \(2012\)](#) que refiere a la arquitectura empresarial como una descripción completa de los elementos de una organización, es decir, sus “partes móviles”, incluidos los objetivos empresariales, las funciones empresariales, los procesos empresariales, la organización en sí misma, las funciones de sus empleados y divisiones, los sistemas de información empresarial, las aplicaciones de software y los sistemas informáticos. Al analizar la tabla encontramos que las empresas del sector se encuentran distribuidas en una sola área, principalmente compañías operadoras (Multinacionales). Se excluye de este estudio el caso de la estatal de Arabia Saudita, ya que siendo consistentes al momento de escribir este artículo, hemos limitado el análisis a empresas con un valor de mercado conocido, que tengan participación abierta (por acciones).

Al ver la orientación de las empresas listadas en particular, y de la industria minero energética en general, vemos que hemos sido testigos de un argumento convincente durante la última década para adoptar técnicas de computación blandas, a medida que los problemas en la cadena de exploración y producción se tornan más complejas para confiar al juicio humano orientado con disciplinas aisladas, emergen así métodos de análisis deterministas y de interpretación integrados. Vemos también una tendencia creciente de

las compañías a vincular el análisis de riesgos y la cobertura de los mismos como una herramienta de mucho peso en la toma de decisión al priorizar su portafolio de proyectos. Nos encontramos en medio de una avalancha de datos a lo largo de la cadena de valor de exploración, producción y comercialización de sus productos que está transformando los modelos basados en datos de una curiosidad profesional en un imperativo de la industria y del sector. ([Holdaway e Irving, 2014](#))

Analizaremos brevemente la estrategia de *big data* de la minera anglo-australiana ([Rio Tinto, 2019](#)) para contextualizar lo que hemos venido diciendo. La empresa creó una unidad de ciencia de datos en 2016 para fortalecer el trabajo con big data que llevaba ya varios años. Rio Tinto ha estado enfocada en producir análisis predictivos para generar información sobre sus operaciones y encontrar más ganancias de productividad que le permitan extraer mayores volúmenes de material mineral o reducir los activos que necesita para mantener los niveles actuales de explotación minera. Rio Tinto ha mantenido una estrecha relación con las universidades de Queensland, Western Australia y el Imperial College of London para formar líderes académicos que promuevan los cambios de competencias que requiere el cambio de cultura organizacional, y para soportar el desarrollo de trabajos de investigación que den soporte a las herramientas de transformación al interior de la empresa.

De esta manera la minera consolida su estrategia “mina del futuro” donde se trabaja en formas avanzadas de extraer minerales de las profundidades de la tierra, mientras se reduce al mismo tiempo los impactos ambientales y se mejora la seguridad de las mismas.

Para ello se combinan análisis de datos, sistemas empresariales, desarrollo de habilidades internas y más de una década de experiencia en el programa para ofrecer a sus accionistas y grupos de interés la promesa del aumento de productividad requerido para garantizar la sostenibilidad de sus operaciones futuras.

En este sentido, para continuar con el ejemplo de la minera Rio Tinto, ahora se centran en mejorar lo que ellos denominan la “plataforma de datos de Rio Tinto”, en la cual tiene mucho valor la simplificación del entorno operativo (variedad), construyendo una plataforma de datos integrada donde puedan visualizar múltiples datos de múltiples orígenes. Es necesario resaltar que, a diferencia de lo que se observa entre las empresas del sector del petróleo y gas, Rio Tinto ha desarrollado una estrategia “agnóstica hacia las plataformas”, es decir, no se ha comprometido con ningún proveedor o tecnología específica. Esta estrategia le ha permitido conectar y reproducir diferentes aplicaciones, ya sean desarrolladas por ellos mismos o por terceros. El valor de traer a mención la estrategia de Rio Tinto viene dado

en que puede ser el derrotero a seguir para asociaciones con quienes se pueda trabajar en conjunto para añadir valor.

Se muestran dos casos de aplicaciones de dominio público de automatización obtenidos por aplicaciones de *big data* en la minera Rio Tinto:

1. Sistema de perforación automatizado: La primera prueba exitosa que demostró la factibilidad de un Sistema Autónomo de Perforación se llevó a cabo en 2008. En 2018, los 11 simulacros habilitados para el sistema habían perforado más de 5.000 kilómetros. El sistema automatizado de perforación permite que un operador que utiliza una sola consola desde un lugar remoto pueda operar múltiples equipos de perforación de múltiples fabricantes.
2. Automatización de trenes para transportar el mineral de hierro a las instalaciones portuarias: los trenes comenzaron a circular en modo autónomo en el primer trimestre de 2017, y más del 60 por ciento de todos los kilómetros de trenes ya se han completado en modo autónomo con un conductor a bordo para su supervisión. Una vez que esté completamente operativo, permitirá importantes beneficios de seguridad y productividad como resultado de la reducción de la variabilidad y el aumento

de la velocidad en toda la red, ayudando a reducir los tiempos de ciclo promedio.

Hasta aquí hemos hablado de cómo las aplicaciones de *big data* han conducido a soluciones de optimización de operaciones, dando el caso particular de una empresa del sector minero. Hemos podido demostrar cómo la orientación al análisis de datos ha traído consigo el desarrollo de nuevas habilidades y la creación de una nueva cultura organizativa que dé lugar a grupos transversales en vez de grupos especializados. Continuando con la misma línea analizaremos brevemente la disrupción que *big data* ha producido en la industria de Petróleo y Gas.

Para empezar, en el centro de las metodologías analíticas multidisciplinarias en la industria de hidrocarburos se encuentran las técnicas de minería de datos, las cuales proporcionan modelos descriptivos y predictivos para complementar el análisis de ingeniería convencional basado en los principios físicos. Los avances en la agregación de datos, la integración, la cuantificación de las incertidumbres y los métodos de computación flexible están permitiendo a la industria de petróleo y gas en su conjunto, perspectivas suplementarias sobre los datos dispares recaudados en las fases iniciales de sus proyectos para crear, por ejemplo, modelos de yacimientos más precisos que pueden ser corridos y ajustados de una manera

más rápida. La computación suave (*soft computing*) es amena, eficiente y robusta, además, requiere menos recursos que la interpretación tradicional basada en las matemáticas, la física y la experiencia de los expertos especializados.

Al interior de cualquier organización o sector, cierto nivel de adopción de cultura tecnológica no es algo que se consiga fácilmente, se debe trabajar para integrar a las personas con las aplicaciones tecnológicas. Siendo claros, la arquitectura organizacional debe transformarse para hacer posible una adopción rápida de las nuevas tecnologías y de la cultura del *big data* en particular. No es el objetivo del presente trabajo analizar los modelos organizacionales ni detallar la incidencia de la cultura empresarial en la rapidez de adopción a las nuevas condiciones, o en este caso particular, a las innovaciones disruptivas como ya hemos visto, pero sí podemos decir que la cultura de cada organización se compone de varias subculturas basadas en el pasado y las suposiciones actuales. Una de ellas es la forma en que las personas trabajan y piensan sobre la tecnología, las TIC y la automatización; la relación de trabajo con las TIC y los sistemas de control se remonta a muchos años, o incluso décadas, para la mayoría de las empresas.

Es posible que se hayan creado muchas suposiciones sobre las TIC; es allí donde las empresas deben definir su cultura organizativa en torno a la

adopción de nuevas tecnologías, nuevas metodologías y nuevos procesos, entendiendo que, en el contexto actual, son procesos de gran dinamismo que requieren exigencia y forzar los ritmos de la respuesta de las organizaciones. En este sentido un auto examen consciente de parte de las organizaciones es un buen comienzo para determinar su velocidad de adaptación a las necesidades y cambios de su entorno.

En el caso del sector de hidrocarburos, es de todos conocido, que a medida que hacemos la transición hacia la adopción de un mayor número de tecnologías y sensores en las operaciones rutinarias, la industria se enfrenta al dilema de cómo adoptar modelos de campos petrolíferos digitales (DOF por sus siglas en inglés); a la par de preguntarse cómo obtener el mayor provecho de los pozos inteligentes con múltiples sensores y medidores. Independiente de la decisión que adopte la compañía respecto a la forma en que conduce sus operaciones, es indiscutible que se está generando a alta velocidad una plétora de datos dispares que definen un paisaje complejo y heterogéneo como un sistema integrado de instalaciones de pozos de almacenamiento. Estos datos de alta dimensionalidad se complementan con datos no estructurados procedentes de la actividad de los medios sociales, y con los dispositivos móviles, demostrando ser valiosos en las operaciones de campo y la computación en la nube, proporcionando una mayor flexibilidad y un mayor rendimiento en la gestión

de redes y datos. Estamos, pues, en una posición ideal para combinar las metodologías de computación de software con los enfoques deterministas e interpretativos tradicionales. (Holdaway e Irving, 2014).

Comentarios finales – analítica del aprendizaje

Big data está creando una gran revolución en muchas áreas de nuestra vida diaria y al mismo tiempo está generando nuevas consideraciones a nivel de valores y acciones conductuales. Las nuevas tecnologías están permitiendo el compromiso, la interacción y la participación de un mayor número de personas. Uno de los resultados de esta creciente presencia de la tecnología de datos de gran tamaño es que las operaciones comerciales están cambiando y aumentando la cantidad de información que generan tan rápidamente que el fenómeno de los datos de gran tamaño está empezando a plantear cuestiones éticas (Davis, 2012), es por ello que cualquier discusión sobre el uso de grandes datos en cualquier industria o mercado suscita preocupación sobre el manejo de la información y su propósito.

Mientras que *big data* ofrece la capacidad de conectar información e innovar nuevos productos y servicios tanto para el beneficio como para el bien social, como toda tecnología, es éticamente neutral. Esto significa que no viene con una perspectiva incorporada

sobre lo que está bien o mal o lo que está bien o mal al usarlo. Las tecnologías *big data* no tienen un marco de valores. Los individuos, los centros de educación y las corporaciones, sin embargo, tienen sistemas de valores, y es solo preguntando y buscando respuestas a preguntas éticas que podemos asegurar que los grandes datos se utilicen de una manera que se alinee con esos valores: “Un marco para la ética de los datos” (Davis, 2012).

Al tiempo que nuestros sistemas operativos se hacen más vulnerables de ser hackeados, los gobiernos, las industrias, los mercados y los individuos en general enfrentan grandes oportunidades, retos, desafíos y amenazas que se dinamizan a un ritmo mucho mayor del que pudiéramos siquiera imaginar.

En este sentido la pregunta ¿cómo pueden las técnicas de aprendizaje de datos analíticos ser una fuente de mejora en la calidad de la educación para enfrentar los desafíos de la innovación, la competencia y la productividad en el siglo XXI? La dejamos de presente para que usted amigo lector saque sus propias conclusiones.

La ciencia y la ingeniería siempre han tenido un impacto masivo en la salud y el bienestar humano. En el mundo moderno, los científicos y los ingenieros tienen la capacidad de causar un mayor impacto (tanto positivo como negativo) en la sociedad (local, nacional

y global) de una manera y a un nivel nunca antes considerado en gran parte por la interconexión.

Desafortunadamente, el público solo se da cuenta de la magnitud de este impacto cuando ocurren los desastres. El análisis y procesamiento de datos, como el diseño en ingeniería y la toma de decisiones en ciencia deben abordar las áreas grises del razonamiento humano, pero la toma de decisiones con soportes éticos no puede basarse únicamente en los valores incorporados en el razonamiento de lo conocido. Si bien es necesario aprender la ética profesional, las normas éticas pueden mejorarse mediante el uso de las modernas tecnologías digitales de vigilancia y control, como ha ocurrido en muchas industrias para avanzar hacia la ética del aprendizaje.

En el mundo de los negocios de hoy, cada función de negocio cuestiona la necesidad de análisis. Los nuevos profesionales de la ciencia y la ingeniería serán responsables de descubrir información a partir de cantidades masivas de datos estructurados y no estructurados para ayudar a dar forma o satisfacer necesidades y objetivos empresariales específicos. El papel del profesional de los datos será cada vez más importante, ya que los diferentes actores se apoyarán cada día más en el análisis de datos para impulsar la toma de decisiones y confiarán más en la automatización y el *machine learning* como componentes centrales de sus estrategias para agregar valor.

La pregunta que debemos hacernos es la siguiente: ¿están los centros de formación y las universidades creando talento analítico, respondiendo y adaptándose a las necesidades de la sociedad moderna interdependiente basada en la información?

La creación de talento analítico no es un esfuerzo único, sino un proceso de continuo ensayo y error. Este talento necesita ser preparado y nutrido en un ambiente donde el arte y la ciencia puedan existir en armonía. La toma de decisiones basada en datos es un viaje, y sin el talento adecuado en todos los niveles académicos y profesionales los beneficios de las ciencias de datos no pueden ser realmente aprovechados. Si queremos potenciar los beneficios de las técnicas de *big data* y el análisis de datos, lo que se requiere es un enfoque interdisciplinario, que se base en las matemáticas, los negocios, la tecnología y las ciencias del comportamiento, y una cultura que alimente el talento e impulse la innovación. En palabras simples necesitaremos formar una mayor cantidad de generalistas y un número menor de especialistas en las aulas de clase y en las empresas.

Al final, la adopción de un enfoque de análisis de datos interdependiente en los diferentes programas de ciencia e ingeniería de las universidades se reduce a la interacción entre las personas. Acá volvemos sobre el trabajo de (Minelli et al, 2013) para concluir que las mejores intenciones pueden salir mal si los

objetivos de los académicos analíticos y los consumidores finales no están alineados. Esto es aún más crítico dada la naturaleza altamente iterativa del análisis que requiere que los generadores y los consumidores de datos trabajen en estrecha colaboración de forma continua. Las actividades principales que deben realizarse para garantizar el éxito de los programas universitarios que responden a estas realidades serán (Minelli et al, 2013):

- Alineamiento organizacional: las universidades deben definir los sectores prioritarios y los problemas que deben resolverse, así como los objetivos que tienen un límite de tiempo, pero que a su vez son medibles y alcanzables. Por muy intuitivo que suene, sin enfoque y dirección, ningún proceso o tecnología hará la diferencia.
- Respaldo y patrocinio del gobierno y del sector productivo; es importante que el equipo de liderazgo de la universidad apoye la toma de decisiones basada en hechos e identifique a los consumidores de datos.
- Identificar a los consumidores también requiere que los académicos puedan socializar ideas y maniobrar la estrategia en varios grupos de negocios. Sin un enfoque consciente en estas actividades, los informes visualmente ricos o los modelos estadísticos sofisticados no

generarán ningún valor comercial real.

- Inversión en capital humano analítico; las universidades y centros de formación necesitan movilizar los recursos necesarios para el análisis de datos, contratar el talento adecuado y retenerlo.
- Existe una creciente demanda de analistas que puedan aprender nuevas habilidades según lo exija la situación, ya sea en

matemáticas, ciencias, negocios o tecnología.

Reconocimientos

Los autores desean expresar su agradecimiento a la cátedra de perforación de pozos de las Universidad Técnica Estatal de Ukhtá (UGTU) y Universidad de Minas de San Petersburgo (SPMI) por permitir la publicación del presente artículo.

Referencias Bibliográficas

- Blystone, D. (2019) Investopedia, Understanding the Alibaba´s business model, Recuperado de <https://www.investopedia.com/articles/investing/062315/understanding-alibabas-business-model.asp>
- Bughin, J. Manyika, J. (2016). Mckinsey & Company, Mckinsey Global Institute. *The Age of Analytics: Competing in a data driven world.*
- Camins, A. (2015). What´s the purpose of education in the 21st century? Recuperado de https://www.washingtonpost.com/news/answer-sheet/wp/2015/02/12/whats-the-purpose-of-education-in-the-21st-century/?noredirect=on&utm_term=.98dbb0038f6e
- Christensen C. (1997). *The innovator´s dilemma: When new technologies cause great firms to fail.* Boston, MA: Harvard Business School Press.
- Davis, K. (2012) *Balancing risk and innovation -Ethics of Big Data.* Sebastopol, CA: O´really media Inc.
- Desjardins, J. (2015). Order from Chaos: *How Big Data will change the world.* Recuperado de <https://www.visualcapitalist.com/order-from-chaos-how-big-data-will-change-the-world/>
- Holdaway, K., Irving, D. (2014). *Harness oil and gas: big data with analytics- Optimize exploration and production with data-driven models.* New Jersey, NJ: Willey, SAS Institute Inc.

- Holland, D. (2012) *Exploiting the Digital Oilfield: 15 requirements for business value*. Houston, TX: Xlibris Corporation.
- International Monetary Fund. (2019) *The World Economic Outlook database* (WEO). Recuperado de [https://www.imf.org/external/pubs/ft/weo/2019/01/weodata/index.aspx#:~:text=The%20World%20Economic%20Outlook%20\(WEO, and%20in%20many%20individual%20countries](https://www.imf.org/external/pubs/ft/weo/2019/01/weodata/index.aspx#:~:text=The%20World%20Economic%20Outlook%20(WEO, and%20in%20many%20individual%20countries)
- Minelli, M., Chambers, M. y Dhiraj A. (2013). *Big Data Analytics. Emerging business intelligence and analytic trend for today's business*. New Jersey, NJ: John Wiley & Sons, Inc.
- Pettey, C. (2011). *Gartner, Press Release*. Recuperado de Gartner Says Worldwide Enterprise IT Spending to reach \$2.7 Trillion in 2012. <http://gartner.com/newsroom/id/1824919>
- Ravindra, S. (2017) *Big Data's potential for disruptive innovation*. Recuperado de <http://dataconomy.com/2017/07/big-data-disruptive-innovation/>
- Rijmenam, M. (2013). *Four ways Big Data will revolutionize education*, Recuperado de <https://datafloq.com/read/big-data-will-revolutionize-learning/206>
- Rio Tinto. (2019). *The Pilbara*, Western Australia. Recuperado de <http://www.riotinto.com/australia/pilbara/mine-of-the-future-9603.aspx>
- Statista. (2019). *The 100 largest companies in the world by market value in 2019*. Recuperado de <https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value/>