



Investigación en educación médica

ISSN: 2007-865X

ISSN: 2007-5057

Universidad Nacional Autónoma de México, Facultad de Medicina

Peñaloza Báez, Marcela Juliana

Big data y analítica del aprendizaje en aplicaciones de salud y educación médica
Investigación en educación médica, vol. 7, núm. 25, 2018, Enero-Marzo, pp. 61-66

Universidad Nacional Autónoma de México, Facultad de Medicina

DOI: 10.1016/j.riem.2017.11.003

Disponible en: <http://www.redalyc.org/articulo.oa?id=349759821009>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UNAM  redalyc.org

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



Investigación en
Educación Médica

<http://riem.facmed.unam.mx>



ARTÍCULO DE REVISIÓN

Big data y analítica del aprendizaje en aplicaciones de salud y educación médica

Marcela Juliana Peñaloza Báez*

Dirección General de Cómputo y de Tecnologías de Información y Comunicación, Universidad Nacional Autónoma de México, Ciudad de México, México

Recibido el 15 de septiembre de 2017; aceptado el 15 de noviembre de 2017

PALABRAS CLAVE

Big data;
Analítica del
aprendizaje;
Procesamiento de
datos automatizado;
Análisis de datos;
Toma de decisiones

Resumen *Big data* y la analítica del aprendizaje son herramientas que permiten aprovechar enormes cantidades de datos que hoy existen en formato digital, y que se generan durante la prestación de servicios de salud y cuando se incorporan plataformas digitales al proceso de enseñanza-aprendizaje en la formación de médicos. El objetivo es diseñar proyectos que exploten las ventajas de las capacidades técnicas actuales para el procesamiento automatizado de datos en grandes volúmenes y distintos formatos, para buscar su entendimiento y mejorar la toma de decisiones. Las asociaciones, los patrones y las tendencias que pueden emerger del análisis de datos generados en la educación médica pueden facilitar el seguimiento e incluso modelar la conducta de los estudiantes, y perfilar su desempeño. Con *big data* y la analítica del aprendizaje es posible transformar grandes cantidades de datos en información de muy alta calidad. Lo anterior es un área que puede ser aprovechada por el educador médico, quien conoce la información que tiene disponible y la que puede obtener mediante alianzas y acuerdos legales, a partir de un problema que se requiera resolver; se necesita apoyo de expertos técnicos pero, en todo momento, el rumbo del proyecto le corresponde a quien ha definido el problema, pues es quien conoce el marco de referencia y las posibles acciones a tomar con los hallazgos y las ideas que surjan. Ciertamente se requiere infraestructura, pero con las posibilidades para el trabajo en la nube, y los recursos disponibles en las instituciones médicas y de educación superior, se puede partir de un primer nivel de *hardware* y *software*, y escalar a plataformas más sofisticadas cuando la combinación de hallazgos en los datos y de los volúmenes a procesar lo justifique. Sin duda, *big data* y analítica del aprendizaje son herramientas en las que el educador médico debe poner atención.

© 2017 Universidad Nacional Autónoma de México, Facultad de Medicina. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Planta Baja Edificio IIMAS, Ala Norte, Circuito Interior, Ciudad Universitaria, 04510, Coyoacán, Ciudad de México, Tel.: +52-55-5622-3665.
Correo electrónico: marcelap@unam.mx
La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

<https://doi.org/10.1016/j.riem.2017.11.003>

2007-5057/© 2017 Universidad Nacional Autónoma de México, Facultad de Medicina. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Cómo citar este artículo: Peñaloza Báez MJ. *Big data* y analítica del aprendizaje en aplicaciones de salud y educación médica. Inv Ed Med. 2017. <https://doi.org/10.1016/j.riem.2017.11.003>

KEYWORDS

Big Data;
Learning Analytics;
Automatic data
processing;
Data analysis;
Decision making

Big Data and Learning Analytics in health and medical education applications

Abstract Big Data and Learning Analytics are tools that allow to take advantage of enormous amounts of data that exist in digital format, and that are generated during the provision of health services and when digital platforms are incorporated into the teaching-learning process in the training of doctors. The objective is to design projects that exploit the advantages of current technical capabilities for the automated processing of data in large volumes and different formats, to seek their understanding and improve decision making. The associations, patterns and trends that can emerge from the analysis of data generated in medical education can facilitate the tracking and even modeling of student behavior, and shape its performance. With Big Data and Learning Analytics it is possible to transform large amounts of data into very high quality information. The above is an area that can be exploited by the medical educator, who knows the information that is available and what can be obtained through alliances and legal agreements, based on a problem that needs to be solved; technical support is needed but at all times, the direction of the project corresponds to the professional who has defined the problem, because he/she knows the frame of reference and the possible actions to take with the findings and ideas that arise. Certainly infrastructure is required but with the possibilities for working in the cloud and the resources available in medical institutions and higher education, projects can start from a first level of hardware and software, and scale to more sophisticated platforms when the combination of findings in the data and the volumes to be processed justify it. Without a doubt, Big Data and Learning Analytics are tools in which the medical educator should pay attention.

© 2017 Universidad Nacional Autónoma de México, Facultad de Medicina. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

A mediados del siglo XIX, el médico inglés John Snow, precursor de la epidemiología, probó con el apoyo de métodos cartográficos que el cólera que se había extendido por Londres era causado por el consumo de aguas contaminadas con materias fecales de un pozo en Broad Street, y recomendó clausurarlo, con lo que disminuyeron los casos de la enfermedad. Con la tecnología disponible hoy, Snow podría haber hecho cruces entre la misma información —las ubicaciones de los casos reportados y los datos de los pacientes afectados— y haber resuelto el problema en cuestión de horas¹. En la actualidad, mucho se ha invertido para lograr el registro digital de los expedientes médicos pero, como reconocen distintos especialistas, es importante equiparar esas inversiones con las herramientas que permiten aprovechar toda esa información². Una de esas herramientas es *big data*.

Big data

Big data es un término que escuchamos cada vez con mayor frecuencia. Con las enormes cantidades de datos que existen en formato digital, y que generamos constantemente con actos cotidianos como el uso de dispositivos electrónicos conectados a Internet, la interacción en redes sociales, el trabajo colaborativo en sistemas digitales, el uso de plataformas para el aprendizaje, entre muchos otros, sumado al incremento de las capacidades de procesamiento automatizado a costos cada vez más bajos, era natural que, en la evolución de los usos y aplicaciones de las tecnologías de información y comunicación en el siglo XXI, se desarrollaran métodos y herramientas para aprovechar esos datos, buscar su entendimiento y mejorar la toma de decisiones.

No se trata solamente del gran volumen de información de la enorme variedad de datos que se pueden procesar, generados tanto por intervención humana como por la comunicación entre las propias computadoras. El entusiasmo radica principalmente en el impacto potencial de los hallazgos que pueden emerger de su análisis, como por ejemplo: incrementos en ventas, reducción de costos, mejoras en la eficiencia de las operaciones, entre muchos otros. *Big data* es la razón por la que empresas como Google, Facebook o Amazon nos hacen llegar información que se ajusta en gran medida a nuestras necesidades potenciales.

La educación médica mediada por computadora es un campo idóneo para la aplicación de *big data*, ya que las plataformas de aprendizaje recopilan datos que permiten dar seguimiento e incluso modelar la conducta de los estudiantes, y perfilar su desempeño.

Se puede definir *big data* como las acciones que se pueden realizar sobre conjuntos de datos a gran escala, cuya variedad y volumen excede la capacidad del *software* tradicional para su captación, gestión y procesamiento en tiempos razonables de acuerdo con la gran velocidad a la que se producen tanto por intervención humana como por la interacción entre dispositivos digitales, con la finalidad de analizarlos, entenderlos, generar hipótesis, tomar decisiones, extraer nuevas ideas y conocimientos, o crear nuevas formas de valor.

Big data se enfoca en encontrar asociaciones, patrones y tendencias entre los datos, a diferencia de otras técnicas que buscan encontrar relaciones causa-efecto o proyecciones basadas en la probabilidad. Por ello, comprende y requiere técnicas, algoritmos y enfoques analíticos en conjunto con nuevas propuestas para diseñar la arquitectura de la información que se procesa.

Esto es, la aplicación de *big data* permite transformar grandes cantidades de datos en información de muy alta calidad, para lo cual se requieren herramientas tecnológicas para la captura, el procesamiento, el almacenamiento, el análisis y la visualización de datos, pero también, y de manera muy importante, se necesita conocimiento profundo de la o las disciplinas relacionadas con la naturaleza de los conjuntos de datos que se están transformando. Muchos especialistas denominan a esto la «ciencia de datos», que el *National Consortium for Data Science* de Estados Unidos, una agrupación de líderes de la academia, industria y gobierno para analizar el conjunto de oportunidades en torno a la ciencia de datos, define como «el estudio sistemático de la organización y el uso de datos digitales para acelerar descubrimientos, mejorar los procesos de toma de decisiones críticas, y habilitar una economía orientada a los datos»³.

En medicina, la tecnología de *big data* puede ser aplicada en la toma de decisiones clínicas, en el seguimiento de enfermedades, en la salud pública y en la investigación. La despersonalización de los datos que se utilicen en estas aplicaciones, es decir, la operación de procesos técnicos que aseguren que la información que se va a procesar proteja la privacidad de los individuos, es parte de los retos éticos que demanda el uso de *big data* en el campo médico.

Analítica del aprendizaje

La inclusión de las tecnologías de información y comunicación en la educación médica posibilita la aplicación de la analítica del aprendizaje, al ampliar el espacio de enseñanza-aprendizaje a través del uso de plataformas digitales como los LMS (*Learning Management System*, sistemas de gestión del aprendizaje), la implementación de exámenes electrónicos, y la interacción digital —sincrónica y asincrónica— entre profesores y alumnos, no solo en aulas virtuales, sino aun durante las prácticas clínicas, tanto a nivel individual como grupal. Lo anterior permite disponer de una gran cantidad de insumos cuyo análisis ofrecerá elementos fundamentados en información para descubrir relaciones y asociaciones entre los datos que permitan proponer innovaciones en la educación médica, entre otras: estrategias de personalización del aprendizaje, seguimiento del progreso académico de los estudiantes y aun la proyección de su rendimiento a futuro; pero también apoya: la valoración del impacto del trabajo docente, la calibración de reactivos para la integración de instrumentos de evaluación, y las mejoras del diseño instruccional.

La analítica del aprendizaje «es la medida, colección, análisis y reporte de datos sobre estudiantes y su contexto con el propósito de entender y optimizar el aprendizaje y el ambiente en el cual ocurre», por lo que emplea los datos de los educandos para «crear mejores pedagogías, centrarse en los estudiantes con problemas y evaluar si los programas diseñados son efectivos y han de mantenerse»⁴. También se define como «el uso de técnicas de análisis para ayudar a enfocar los recursos instruccionales, curriculares y de apoyo para impulsar el logro de objetivos de aprendizaje específicos»⁵.

Las aulas virtuales en sistemas como Moodle o Blackboard, y los repositorios digitales que se asocian a estas,

almacenan grandes cantidades de datos al registrar las huellas de la interacción entre los estudiantes y los docentes, y aun entre los propios alumnos. Para convertirse en analítica del aprendizaje, se requiere que esos datos sean procesados para identificar patrones de conducta, tanto de los educandos como de los educadores médicos, y de esta manera contar con elementos objetivos para evaluar el impacto y los resultados de la aplicación de iniciativas institucionales para la mejora de los programas de estudio y los métodos pedagógicos asociados a ellos.

Tipos de datos de *big data*

Big data se basa en la posibilidad de procesar enormes cantidades de datos, pero a diferencia de otras tecnologías, no se requiere que la información esté totalmente estructurada. Es decir, se pueden emplear datos aun cuando estos no tengan el mismo formato (como puede ocurrir si se tienen archivos jpg, png, mp3 o amr), o que sean de muy diferentes tipos (videos, audios, textos, imágenes, entre otros). Las herramientas tecnológicas que habilitan *big data* permiten su análisis de manera ágil y mucho menos costosa que lo que tomaría «normalizar» esos datos, lo que implicaría transformarlos en información estructurada para su procesamiento con tecnologías tradicionales como las bases de datos relacionales.

Aunque *big data* permite procesar datos no estructurados, por supuesto una de las primeras fuentes de información que alimentan sus aplicaciones son, precisamente, los datos de los sistemas con bases de datos tradicionales, entre ellos: aplicaciones administrativas como la contabilidad y la nómina, las operaciones de compra-venta, los directorios de clientes y el historial de sus operaciones, los expedientes clínicos, y todas las transacciones a sistemas de información que se realizan en línea y desde dispositivos móviles.

Existen otro tipo de sistemas en las organizaciones que también aportan insumos de valor para las aplicaciones de *big data*, ya que su operación implica generar datos constantemente, los cuales se almacenan por ser la memoria de su funcionamiento; entre ellos se encuentran las bitácoras de las visitas recibidas en los sitios web o de las llamadas realizadas desde un *call center*, las mediciones de aparatos, y los registros de equipos como veletas y anemómetros, dispositivos móviles y sensores industriales; todo lo que pueda medir, registrar y comunicar el posicionamiento, la vibración, la temperatura, la humedad e incluso los cambios químicos del ambiente. Se estima que hay más de 30 millones de sensores interconectados en distintos sectores como automotriz, transportación, industrial, servicios y comercial, y se espera que este número crezca en un 30% anualmente⁶.

Con *big data* se pueden aprovechar las bases de datos de encuestas y estudios, con información tanto estructurada como no estructurada, y aun con insumos del sector público, como los censos de población, los registros fiscales, las videograbaciones del tráfico, entre muchos otros.

Las plataformas digitales como los blogs y las redes sociales son una fuente importante de datos para las aplicaciones de *big data*, y presentan retos similares a los que implica el procesamiento de otros tipos de insumos como los videos, los audios, mapas y datos de ubicación de los GPS. Cada

visita, cada clic, cada inicio de sesión y cada comentario son almacenados en algún lugar. Además de la estimación global de que existen más de 1,000 millones de páginas web en Internet, estadísticas de Brandwatch⁷, un servicio para el análisis de redes sociales, señalan que: Facebook Messenger y Whatsapp manejan 60,000 millones de mensajes diarios, a través de WordPress se publican mensualmente 56 millones de entradas en sus blogs, Facebook genera 4 nuevos petabytes de datos cada día (como referencia, si partimos de los gigabytes que almacenamos en memorias USB, un petabyte es equivalente a un millón de gigabytes, o bien a 1,000 terabytes); cada minuto se suben 300 horas de video a YouTube, 80 millones de fotos son compartidas en un día promedio en Instagram, Google procesa 40,000 búsquedas por segundo y todos los días se envían 500 millones de mensajes a través de Twitter.

Existen algunas características que distinguen los datos que pueden ser explotados con técnicas de *big data* para aplicaciones en la educación. Prakash et al. señalan que los datos en educación son jerárquicos: al nivel de la interacción con los dispositivos (como los clics o los *likes*), al nivel de las sesiones de los usuarios, al nivel de los alumnos y la interacción entre ellos, al nivel de cada grupo o salón, del maestro y de ahí en los niveles que puedan seguir, de acuerdo con la organización de cada institución educativa: departamentos académicos, clínicas, escuelas o facultades, e institutos o universidades; todos los niveles están «anidados» uno dentro del otro. Otras características a analizar son: el contexto, para determinar bajo qué condiciones funciona un determinado modelo; el tiempo, que indica por ejemplo cuánto invierte el alumno o en qué momentos del día le es más viable interactuar con la plataforma de aprendizaje; y la secuencia, para evaluar el orden que se eligió en el diseño instruccional para la presentación de conocimiento y la realización de actividades⁸.

Aplicaciones en la educación médica

Entre las fuentes de *big data* disponibles en medicina se encuentran: expedientes y registros clínicos y administrativos, datos biométricos, información reportada por el paciente, imágenes médicas, datos de biomarcadores, estudios prospectivos y ensayos clínicos⁹. Los hospitales, los laboratorios, las compañías de seguros y la industria farmacéutica pueden contribuir con grandes cantidades de datos financieros, de enfermedades y de medicamentos¹⁰. De la naturaleza de estos datos se desprenden distintos retos para el aprovechamiento de *big data* en medicina, tales como la disposición a compartirlos y los acuerdos legales y éticos adecuados para hacerlo, el gran volumen de insumos de información con el que se contará, la despersonalización de los datos privados de los pacientes, la disparidad y aun la periodicidad de las fuentes, ya que algunas aportarán datos cada segundo, y otras se actualizarán de manera anual. También es importante considerar los desafíos para la recolección de información por autores tan diversos como personal médico y de enfermería, administradores, alumnos, encuestadores, personal de compañías de seguro y farmacéuticas, entre muchos otros, lo que contribuye a la enorme variedad de datos que pueden ser insumos de las aplicaciones de *big data* en medicina.

Además de la diversidad de actores que los recopilan, es necesario considerar los procesos de normalización que permitan codificar la información que se registra en los textos tanto manuscritos como digitales de los reportes médicos, así como la complejidad inherente a que los datos que se generan a través del tiempo durante la evolución de un caso clínico guardan distintos formatos, por ejemplo, imágenes, reclamaciones de seguro y aun comentarios hechos en medios sociales, desde la fase de diagnóstico y hasta el cierre de cada caso.

Todo lo anterior, bajo un marco de legalidad, códigos de ética, transparencia, confidencialidad, protección de datos personales y de su privacidad, para lo cual se requiere consentimiento de los pacientes y los procesos de despersonalización de la información que se han mencionado, donde la diversidad de leyes aplicables en cada país puede ser un factor que detenga la realización de proyectos de alcance internacional.

A partir de datos tan heterogéneos, las aplicaciones de *big data* en medicina pueden promover grandes beneficios, como la prevención de enfermedades, la evaluación de los costos de la atención médica, la valoración de la eficacia de los medicamentos y la planeación de mejoras en las políticas de salud pública.

Los datos que los expertos en educación médica pueden aprovechar para aplicar las técnicas de *big data* en el marco de la analítica del aprendizaje se recopilan a partir de «acciones estudiantiles explícitas, como completar las tareas y tomar exámenes, y de las acciones tácitas, incluyendo las interacciones sociales en línea, actividades extracurriculares, mensajes en los foros de discusión», entre otras⁴. Por su parte, Prakash et al. proponen ocho áreas de aplicación para la analítica del aprendizaje⁸:

- a) Modelos del conocimiento del usuario, que buscan indagar sobre los contenidos y las habilidades que el estudiante domina, con base en sus respuestas, los errores cometidos y repetidos, las habilidades que practicó y su desempeño en general a partir de su trabajo en plataforma o bien por los resultados obtenidos en exámenes estandarizados.
- b) Modelos de la conducta del usuario, con el objetivo de conocer más sobre la motivación de los alumnos y los patrones de interacción que muestran los estudiantes a través del proceso de aprendizaje, por vía de datos muy similares a los del punto anterior, a los que se agrega el análisis de los cambios en el contexto de cada alumno durante un período.
- c) Modelos de la experiencia del usuario, para evaluar su nivel de satisfacción, a través de sus respuestas a encuestas y cuestionarios, y de la evaluación de las decisiones que tomen y el desempeño que muestren en los siguientes cursos o unidades de aprendizaje.
- d) Perfiles de los estudiantes, que se logran al identificar los grupos de alumnos a partir de los datos disponibles que se mencionan en otras áreas de aplicación. Con los perfiles definidos pueden diseñarse acciones remediales, por ejemplo, si se detecta que un grupo ha tenido dificultades en el desarrollo de ciertas competencias.

- e) Modelos de los dominios, para valorar el nivel de desagregación que más conviene para dividir los temas en módulos y la secuencia en que deben ser presentados.
- f) Análisis de los componentes del aprendizaje y de los principios instruccionales, para identificar los elementos que son más eficaces.
- g) Análisis de tendencias, para indagar sobre el impacto de las acciones a través del tiempo, para lo cual se requieren datos sobre al menos tres momentos, con la finalidad de lograr identificar una tendencia; la información requerida incluye registros de ingreso y egreso de alumnos.
- h) Adaptación y personalización, que es uno de los objetivos de la mayoría de los proyectos de analítica de aprendizaje, con el objetivo de identificar las acciones idóneas que se deben sugerir a cada individuo, y cómo debe modificarse la experiencia para el siguiente usuario a partir de lo sucedido previamente, tanto para el propio estudiante como para otros alumnos y, de ser posible, lograr este nivel de personalización en tiempo real. Para ello normalmente se requiere recolectar datos históricos sobre cada usuario, de la misma manera como lo hacen las redes sociales.

Para aplicar los beneficios de *big data* en la educación médica es necesario promover un cambio de paradigma no solo entre alumnos y profesores, sino también entre los profesionales y todos aquellos actores involucrados en el cuidado de la salud, con la finalidad de promover que no exista brecha entre los ambientes en los que se forman los médicos, respecto de la práctica profesional, la prestación de servicios asociados, la educación continua y la investigación médica, pues se espera una fuerte tendencia hacia la personalización del cuidado de la salud¹⁰. Lo anterior requiere asimismo que los responsables de estos programas sean entrenados en la gestión del cambio cultural que todo esto representa.

El programa *Health Care by the Numbers* de la Universidad de Nueva York (NYU) es un ejemplo de la aplicación de *big data* en la educación médica. Se trata de un programa curricular que tiene como objetivo la formación de los estudiantes de medicina en el planteamiento y ejecución de proyectos de *big data* con la finalidad de identificar áreas de oportunidad para la mejora de los cuidados de la salud. Para ello, los alumnos disponen de datos despersonalizados provenientes tanto de los servicios externos del Centro Médico Langone de la propia NYU como de datos abiertos gubernamentales de los más de 2.5 millones de pacientes admitidos cada año en los hospitales estatales (datos que pueden ser utilizados, reutilizados y distribuidos libremente por cualquier persona y que se encuentran sujetos al requerimiento de atribución y de compartirse en los mismos términos en que fueron publicados). La intención es que los futuros médicos adquieran competencias en el manejo de herramientas que les permitan desarrollar programas tanto para el cuidado de la salud individual como de grupos de la población¹¹. Los datos pueden incluir edad, grupo racial, grupo étnico, diagnósticos, procedimientos, costos, entre otros, y el hecho de estar despersonalizados implica que no pueden conocerse los nombres de los pacientes, o datos que conduzcan a su identificación. Con las herramientas de *big data*, los estudiantes pueden hacer pronósticos médicos basados en la información procesada, o incluso proponer sistemas de

farmacovigilancia para rastrear las reacciones adversas a los medicamentos o los errores en la práctica¹⁰. Todo ello, con datos reales.

Ellaway et al. proponen algunas aplicaciones de *big data* en la educación médica, entre ellas: poder determinar cuándo se puede permitir atender pacientes vivos a un médico en formación que ha practicado en simuladores de laboratorio, a partir de los resultados obtenidos en las aplicaciones de computadora. Otra aplicación interesante resulta de combinar datos de acervos de instituciones educativas con repositorios de información clínica, para correlacionar los resultados de los tratamientos a los pacientes que se almacenan en los sistemas de información de los hospitales, con los diferentes modelos educativos de cada escuela de medicina, o de un grupo de instituciones. También se puede pensar en comparar datos de distintas instituciones educativas en el mismo período, en relación con el desempeño de los alumnos, para rastrear el impacto de eventos significativos en el contexto escolar tales como desastres naturales, epidemias, conflictos sociales o crisis políticas¹².

Establecer proyectos similares para enriquecer la educación médica requiere la colaboración de los educadores médicos con los profesionales de las TIC, para la provisión de infraestructura de *hardware* y *software*, y con los equipos legales para el establecimiento de acuerdos para el intercambio de datos. Es necesario también interactuar con los profesionales en ciencia de datos, que son especialistas que al día de hoy se forman a través de programas de posgrado, y que se espera que en el mediano plazo se cuente con oferta de carreras afines para el nivel de licenciatura. Las competencias que se requiere integrar en los equipos de trabajo multidisciplinarios incluyen el análisis predictivo, visual y gráfico, así como el procesamiento de lenguaje natural, la inteligencia artificial y las técnicas para cosechar —recopilar— datos no estructurados.

¿Qué hacer para iniciar un proyecto de *big data* o analítica del aprendizaje?

Para empezar a definir un proyecto de *big data* o analítica del aprendizaje para la salud o la educación médica, el primer paso es hacer un inventario de la información que se tiene disponible, y de aquella que se puede obtener mediante el establecimiento de alianzas o acuerdos locales o aun interinstitucionales. En paralelo, es necesario definir el problema que queremos resolver, o la línea temática en la cual se buscarán las tendencias y los patrones que resulten del análisis de los datos disponibles. Para mejor referencia, se recomienda revisar algunas demostraciones disponibles en línea¹³⁻¹⁵.

Se requerirá el apoyo de expertos técnicos cuando se tengan datos en distintos formatos, por ejemplo: videos, imágenes y publicaciones en redes sociales, pero en todo momento es importante tener presente que el rumbo que tome el análisis y el procesamiento de los datos será establecido por quien esté interesado en resolver el problema que se definió, el cual conoce su marco de referencia y las posibles acciones a tomar con los hallazgos y las ideas que surjan.

Por supuesto se requiere infraestructura, pero con las posibilidades para el trabajo en la nube y los recursos disponibles en las instituciones médicas y de educación superior, se puede contar con elementos técnicos y escalearlos de acuerdo con los volúmenes de datos a procesar y en razón de los resultados que arrojen las primeras corridas de análisis.

Conclusión

Es indudable que el uso y el aprovechamiento de la tecnología es un diferenciador entre los profesionales del cuidado de la salud, por lo que la incorporación de técnicas de *big data* y de analítica del aprendizaje para los estudiantes y profesores de medicina es una necesidad que cada vez tendrá mayor demanda, pues también requiere la interacción con otros especialistas, a fin de lograr sus beneficios. En este artículo se han presentado los principales conceptos en torno a estas herramientas, pero sin duda es importante resaltar que para su aplicación es necesario contar con la visión y el apoyo de las instituciones educativas, por las inversiones, decisiones y cambio cultural que implica la incorporación de proyectos afines en los programas de educación médica.

Los proyectos que se pueden desarrollar en la educación médica con apoyo de *big data* y analítica del aprendizaje dependen en gran medida de los niveles de acceso a los datos disponibles, sus formatos y su calidad, en términos de la precisión, representatividad y confiabilidad que se alcance bajo las consideraciones señaladas en el texto sobre la diversidad de actores que participen en la generación de los insumos de estos proyectos. Se requieren estándares de datos, métodos y procesos técnicos y analíticos sobre ellos, para intentar evitar caer en sesgos que puedan derivarse de no contar con todos los elementos necesarios para el procesamiento de la información disponible. En su momento, y de haber contado con más información de la que se requería procesar, John Snow pudo haber llegado a conclusiones equivocadas respecto a las causas de la epidemia de cólera, de ahí la importancia de contar con métodos robustos y profundo conocimiento de la disciplina.

Aun con los retos mencionados, se considera que el educador médico debe familiarizarse con *big data* y la analítica del aprendizaje para incorporar estas técnicas al conjunto de herramientas con las que se pueden preparar proyectos multidisciplinarios de innovación educativa, con las ventajas que implica poder basar su diseño y la toma de decisiones en una cantidad cada vez mayor y más diversa de datos, en contraste con los métodos de investigación tradicionales.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que en este artículo no aparecen datos de pacientes.

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes.

Financiación

Ninguna.

Conflicto de intereses

La autora declara no tener conflicto de intereses.

Referencias

1. Khoury MJ, Ioannidis JP. Big data meets public health. *Science*. 2014;346:1054–5.
2. Dexler M. Big data's big visionary. *Harvard Public Health Maga-*
3. Ahalt S, Bizon C, Evans J, Erlich Y, Ginsberg G, Krishnamurthy A, et al. Data to discovery: Genomes to health. A White Paper from the National Consortium for Data Science. Chapel Hill (NC): University of North Carolina (US); 2014. p. 1.
4. Navarro V, Cragno A, DiBernardo J. Analítica del Aprendizaje aplicada a los estudiantes de Medicina en Foros de discusión de casos clínicos en Medicina III. *Rev Arg Educ Méd*. 2016;7:14–8.
5. Barneveld AV, Arnold KE, Campbell JP. Analytics in Higher Education: Establishing a common language. *EDUCAUSE. ELI Paper 1*; 2012:1-11.
6. Barranco R. ¿Qué es Big Data? [Internet]. México: IBM Software Group México; 2012. Disponible en: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/que-es-big-data-pdf.pdf>. Consulta: 6 de septiembre de 2017.
7. Smith K. 96 estadísticas y datos increíbles de las redes sociales para 2016 [Internet]. Inglaterra: Brandwatch; 2016. Disponible en: <https://www.brandwatch.com/es/2016/08/96-estadisticas-redes-sociales-2016/>. Consulta: 6 de septiembre de 2017.
8. Prakash BR, Hanumanthappa M, Kavitha V. Big Data in educational data mining and learning analytics. *IJIRCE*. 2014;2:7515–20.
9. Lee CH, Yoon HJ. Medical big data: Promise and challenges. *Kidney Res Clin Pract*. 2017;36:3–11.
10. Lateef F. Big Data: Applications in healthcare and medical education. *Education in Medicine Journal*. 2016;8:85–9.
11. Institute for Innovations in Medical Education. Healthcare by the numbers [Internet]. Estados Unidos: NYU School of Medicine; 2014. Disponible en: <https://med.nyu.edu/institute-innovations-medical-education/research-scholarship/grants/health-care-the-numbers>.
12. Ellaway RH, Pusic MV, Galbraith RM, Cameron T. Developing the role of big data and analytics in health professions education. *Med Teach*. 2014;36:216–22.
13. Stratebi. Demos Big Data Online [Internet]. España: Stratebi; 2017. Disponible en: <http://bigdata.stratebi.com/>.
14. Zoomdata. Zoomdata Product Demos and Trials [Inter-net]. Estados Unidos: Zoomdata; 2017. Disponible en: <https://www.zoomdata.com/free-trial/>.
15. Syncfusion. Showcase samples [Internet]. Estados Unidos: Syncfusion; 2017. Disponible en: <http://datascience.syncfusion.com/bigdata>.