

BIG DATA aplicado a la medicina

A. F. Casallas, & D. A. Jimenez

Abstract— En este artículo se busca determinar el impacto del BIG DATA en el campo de la medicina y nuestra vida cotidiana. Además de dar a conocer varias investigaciones y elementos para fomentar el uso de esta en este campo. Para ello en las siguientes líneas tratamos de aclarar lo que significa el Big Data, de describir su implementación en el mundo de la medicina, y presentar una panorámica, tanto de lo que el Big Data puede aportar en la construcción de una sociedad mejor, como de los problemas que puede generar. Así logrando proporcionar una visión de cómo se encuentra ahora mismo de desarrollado, a qué áreas está enfocado y a qué público va dirigido, puede ayudarnos a predecir comportamientos, a encontrar nuevos patrones y, en definitiva, a indicarnos nuevas maneras de ver el futuro.

Palabras Claves— Medicina, BIG DATA, Minería de datos, información

I. INTRODUCCIÓN

En este artículo nos centraremos en explicar algunos métodos de recolección de datos en la medicina, en cómo se hace la recopilación de datos para diagnosticar que posibles enfermedades tenga el paciente además de enfatizar en las facilidades que ofrece. Pero antes de eso debemos definir qué es el BIG DATA.

El uso de minería de datos o BIG DATA en Estados Unidos en salud puede ahorrar a la industria de la salud hasta 450 mil millones de dólares cada año (Basel Kayyali, David Knott, & Steve Van Kuiken, 2013). Esto se debe a los volúmenes crecientes de datos generados y de las tecnologías para analizarlos.

Los datos que resultan útiles en la investigación en salud provienen de fuentes diseñadas para la investigación como las historias clínicas, pruebas de laboratorio, censo de población, registros de enfermedades, etc. El Big Data o minería de datos, la hace diferente el hecho de incorporación de otras fuentes de información ya sean los obtenidos por medio de Apps (dispositivos que integramos a nuestro cuerpo: celulares, relojes...), las redes sociales o las plataformas "nube", entre otras, y la posibilidad de realizar la consultar aquellas fuentes.

BIG DATA es un término que nació del crecimiento constante de volumen o tamaño de la información. La expresión surgió en el año 2009 y es aplicable a múltiples

campos, termina siendo importante en ámbitos como economía, tecnología o en empresas.

Crecimiento del volumen de datos en el mundo

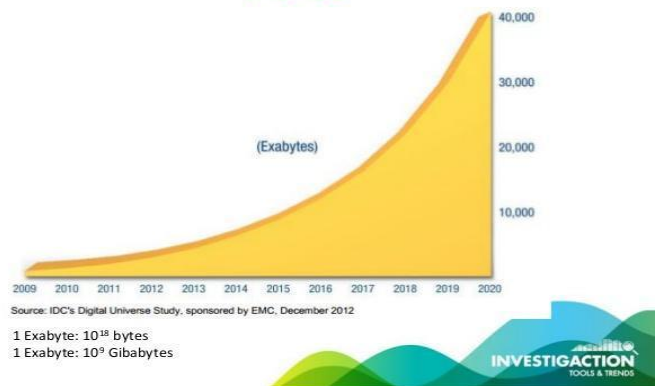


Ilustración 1. Crecimiento del Volumen de Datos en el Mundo.

Los datos provienen de diversas fuentes como transacciones bancarias, pagos con tarjeta e incluso búsquedas en internet, también hay datos que nosotros proporcionamos al publicar algo, subimos imágenes o videos entre otros...

Para poder saber lo grande que es el BIG DATA, empezaremos por el megabyte, que equivale a 1 millón de bytes y nos permitiría almacenar un libro de unas 500 páginas que tuviera sólo texto (Grant, 2012). En un terabyte, equivalente a un millón de megabytes, podríamos almacenar 2.767 copias de la Enciclopedia Británica, ó 16.667 horas de música, ó 1.333 horas de video. Se necesitan casi 5 exabytes, ó 5 millones de terabytes, para almacenar las secuencias del genoma humano de todas las personas del mundo. Pues bien, la compañía americana IDC estima que en 2011 se crearon 1,8 zettabytes de información, es decir, 1.800 exabytes, una cantidad de datos con la que podrían llenarse 57.500 millones de iPads de 32 gigabytes. Con este número de iPads podría construirse una pared dos veces más alta que la muralla china. (Estrella Pulido Cañabate, 2015).

Toda esta cantidad de información se genera a una enorme velocidad, tan alta que el 90% de los datos que existían en el mundo en 2012 se había creado en los dos años anteriores. Cada minuto, por ejemplo, se envían 150 millones de mensajes de correo electrónico, se realizan más de 2 millones de búsquedas en Google y se visualizan casi tres millones de videos en YouTube (Excela, 2016). En media hora, el motor

A. F. Casallas. Universidad Libre de Colombia, Bogotá – Colombia Andresf.casallas@unilivre.edu.co

D. A. Jimenez. Universidad Libre de Colombia, Bogotá – Colombia Daniela.jimenezb@unilivre.edu.co

Corresponding author: D. A. Jimenez

de un avión que vuela de Londres a Nueva York genera 10 terabytes de datos 4 (MacKinnon, 2013).

En el mundo actual el big data abarca muchos campos y juega un papel muy importante. Antes la medicina se había convertido en analizar a un cierto grupo de personas y diagnosticarlas a cada una por aparte. Esto generaba algunos problemas en la población, por ejemplo, las personas que asistían muy tarde al médico cuando tenía una enfermedad latente todo el tiempo. Además de generar sobre costos en la atención de pacientes y en algunos casos citas innecesarias al médico.

Con el BIG DATA se puede tener una información poblacional acerca de los, enfermedades, historial médico, y estado actual del paciente. Esto genera que se pueda predecir una posible epidemia a tiempo mediante un repositorio que genere alertas tempranas, además esta herramienta permite detectar enfermedades a corto plazo y genera una mayor organización en el sistema de salud. Además de generar patrones de comportamiento y afectación respecto a las dolencias de la gente.

En realidad, parece inminente la era del BIG DATA, con el lograremos una amplia mejora de nuestra capacidad para realizar diagnósticos y pronósticos fiables en numerosas áreas de la vida social. Esta nueva capacidad de análisis y predicción puede ser aprovechada por las empresas con fines puramente lucrativos utilizando distintas técnicas, como la publicidad personalizada o el análisis de sentimientos.

El Big Data tiene las características que se aplican a todos los sectores sin que medicina se la excepción conocidas como las cinco "V": Volumen, velocidad, variedad, valor de los datos y la veracidad de datos.

En el campo de la sanidad existen una enorme cantidad de datos entre ellos registros médicos, rayos X, pruebas de laboratorio, imágenes médicas, datos de ensayos clínicos, altas, datos a nivel genético, las secuencias genómicas de datos de población, dispositivo actuales etc. Estos datos deben ser de calidad, esto es vital en lo que se refiere a salud por el motivo de decisiones de tipo vida o muerte, si no es precisa esta información es un problema grave ya que de eso depende una persona. Para evitar problemas como el mencionado anteriormente los datos se clasifican en estructurados y no estructurados, los primeros son aquellos que se pueden almacenar, consultar, analizar y manipular por ordenador fácilmente. Históricamente, en la asistencia sanitaria y los segundos son lo contrario a lo datos estructurados un ejemplo son lo formato multimedia (Merino, 2016).

No obstante, para que los usuarios que forman parte de los diferentes entornos el campo de la salud cumpla su objetivo de poder tomar las mejores decisiones a partir de la información que se tiene, es indispensable también disponer de una infraestructura de Business Intelligence (BI), que integre las herramientas adecuadas para facilitar la extracción y visualización de información de Big Data. Los principales proveedores de este software BI (Oracle, IBM, Microsoft, ...), son conscientes de las limitaciones que había y han

comenzado a incorporar este tipo de herramientas en sus últimas versiones de sus productos.

Cabe aclarar que Business Intelligence (BI), es el proceso mediante se transforman datos en información y esta se transforma en conocimiento, se usa mucho en la toma de decisiones en los negocios.

En la gran cantidad de datos el volumen de cada uno de ellos puede variar, aunque muchos de ellos son de gran tamaño, por ejemplo, un solo genoma de una persona ocupa 3 GB de ahí mismo pueden salir radiografías, resonancias magnéticas, entre otros; por este motivo es difícil almacenarlos solo en una base de datos tanto por su tamaño y velocidad de cómo son analizados. Esto nos lleva a las cinco V del Big Data, estas nos permiten:

Personalización: ofrecer un tratamiento adecuado y específico a cada paciente, también se puede tomar en el sentido de por ejemplo una persona tiene cáncer, pero cada persona su cuerpo recibe cada medicamento de forma diferente debido a su genoma o genes, gracias al Big Data se puede especificar estos casos con más eficiencia e incluso nos lleva a predecir síntomas de enfermedades antes de que ocurran.

Prevención: e incluso nos lleva a predecir síntomas de enfermedades antes de que ocurran, por medio de los genes se pueden detectar distintas propiedades que hay en ellos, una vez detectadas se puede seguir el respectivo tratamiento para evitar el mal.

Participación: este tipo de medicina le permite al paciente ayudarse por medio del internet por donde consigue información sobre su enfermedad.

En la salud la minería de datos también ayuda en temas como la incapacidad laboral, consiste en la situación de pérdida de la capacidad de laboral consecuencia de impedimentos funcionales provenientes de una enfermedad o lesión. Cabe aclarar que existen dos tipos de incapacidad: la incapacidad permanente, tiene un plazo de trescientos sesenta cinco días entendibles por otros ciento ochenta días gracias a que el trabajador se pueda curar y dado de alta por la atención (esto lo dice el artículo 128 TRLGSS (Ley General de Seguridad Social)); incapacidad permanente que es básicamente que la persona ya no se puede curar ya sea de la lesión o enfermedad, mientras el paciente tenga estas limitaciones recibirá tratamiento médico.

¿Pero qué tiene que ver el big data?, simplemente en el hecho de aportar las mejores posibilidades de análisis y relación, basada en la evidencia, optimizando y aportando referencias para la mejor distribución de recursos y la ayuda a la toma de decisiones tanto por rapidez como precisión lo que hace urgente implementar para obtener mejor situación, análisis y planificación estratégica a la hora de realizar una incapacidad.

Para mencionar otro ejemplo una rama de la medicina es la telemedicina, es la atención médica o el intercambio de información médica mediante el uso de las telecomunicaciones o sistemas de información, en pocas palabras, a distancia. Esto incluye no sólo prestación del servicio médico a pacientes (consulta, diagnóstico y

tratamiento), también la enseñanza y el intercambio de datos o conocimientos médicos.

El no encontrarse presencialmente en un lugar para diagnosticar a alguien, el poder monitorizar a un paciente sin que éste se encuentre hospitalizado, el poder aprender y practicar un procedimiento como si el paciente se encontrara ahí, el intercambio de información desde diferentes partes del mundo, todas estas posibilidades son hoy reales gracias a la Telemedicina siendo posible y mejor por el Big Data.

Existen diferentes herramientas dos ejemplos muy conocidos son: WuXi NextCODE Se trata de una compañía que utiliza Big Data en información genómica para ayudar a realizar investigaciones que mejoren la salud de las personas. Cuentan con una plataforma capaz de organizar, analizar y compartir esa información, y es la única que dispone información a nivel de población y no de personas; Y Predicción software, Se trata de un software que se instala en diferentes aparatos médicos, desde ultrasonidos a medidores de glucosa, y que monitoriza en tiempo real el estado de estos. Además, utiliza la tecnología Big Data para analizar ese estado y predecir cualquier anomalía que pueda afectar al estado de la máquina, de esta manera, se permite planificar las intervenciones médicas en función de información real sobre las máquinas disponibles. También permite observar con antelación si existe algún síntoma de problema, pudiendo darle solución incluso antes de que ocurra, y conocer el rendimiento del aparato.

Aun con todas las ventajas existentes aún presenta inconvenientes. En la minería de datos es la exploración automática o semiautomática de los grandes conjuntos de datos con la intención de descubrir patrones, en el cual se incluyen la recolección y la preparación de los datos, la interpretación de los resultados y la información de estos. Sin embargo, genera ciertos desafíos para la ciencia actual. El primero, es la búsqueda de patrones en los big data: para ilustrarlo, Y el segundo es la posibilidad de asociaciones inexistentes aparentemente pero que pueden existir por una tercera relación no vista.

Existen diferentes barreras que representan un problema aún en la actualidad podríamos dividir las en: Barrera administrativa, es necesario que exista un acuerdo entre todas las partes involucradas para llevar a cabo la distribución de información; Barrera tecnológica, por más que hoy en día haya grandes avances tecnológicos aún falta mejorar en lo que se refiere a Big Data, y aún más en medicina, debería adaptarse más para poder dar cada vez análisis más rápidos y exactos; por último la Barrera legal: Al ser el autor principal los datos en el Big Data lo más probable es que gran cantidad de ellos sean personales de cada paciente, con ellos se podría hacer diferentes actos corruptos, esta barrera consiste en evitarlos.

En el último aspecto nos centraremos más ya que barrera legal también se puede entender seguridad El análisis masivo de datos hace ver ciertos riesgos, como el de la privacidad de las personas. Las técnicas de disociación (los datos de una persona se asocian a un código para evitar su identificación, aunque se sabe que el conjunto de datos pertenece a una determinada persona) cobran un papel muy importante en la

de recolección y procesamiento de datos para la creación de modelos predictivos propia de fase del big data

Por ejemplo, La regulación europea prevé la protección de los datos personales, entre los que se encuentran los de la salud de la ciudadanía. Sin embargo, existen países donde la normativa puede ser más laxa o inexistente, y donde obtener estos datos puede ser más fácil. Por otro lado, la dependencia económica de los países de renta baja imposibilita que ejerzan la soberanía sobre sus datos frente a los países de renta alta. A lo anterior se le podría añadir que en la mayoría de los casos es difícil que los resultados de las investigaciones reviertan en la población que los ha originado, debido a la inestabilidad política, la corrupción, la pobreza y la precariedad de los sistemas de salud y del acceso a las nuevas tecnologías. Los avances científicos que se deriven de esas investigaciones deberían mejorar la salud y los determinantes de la salud de la población en esos países (Facundo, 2016)

II. CONCLUSIONES

Se concluye que las buenas prácticas en la investigación en salud pública deben ser diferentes para las investigaciones que usan big data. Por tanto, la división entre la investigación con big data y la investigación tradicional no parece lógica.

Los investigadores e investigadoras de la salud pública deberían desempeñar un papel central en la propuesta de hipótesis innovadoras, en la construcción de infraestructuras para el almacenamiento de grandes conjuntos de datos y en asegurar el desarrollo de enfoques sistemáticos en el análisis de grandes conjuntos de datos complejos y masivos.

El Big Data presente en salud quiere lograr un gran reto, este reto busca optimizar y mejorar condiciones en lo que se refiere a salud tanto a favor del paciente, médico, aprendices, y la comunidad de sanidad. Con tal objetivo propuesto se pueden usar los diferentes datos en mejorar la capacidad de respuesta y la asistencia médica a pacientes crónicos o personas con necesidades socio asistenciales, entre otros; La predicción de reingresos de pacientes, que permitiría una reducción de costos, una mejora de la asistencia sanitaria y un aumento en la calidad del servicio. Estas tareas son posibles gracias a la aplicación de metodologías de minería de datos.

Para alcanzar estos logros es fundamental la integración de todos los datos provenientes de diferentes fuentes, así como el desarrollo de nuevas tecnologías que permitan la explotación de dichos datos. No obstante, el verdadero valor de Big Data en salud se conseguirá sólo si los diferentes actores implicados en el proceso (administraciones públicas, empresas privadas, hospitales, médicos, centros de investigación, universidades, ...) se comprometen en este proyecto de forma conjunta para llevar el ámbito sanitario a una nueva era. Esto solo se podrá llevar a cabo en el marco de un ecosistema de Big Data en salud en el que se integre juntamente con la tecnología, políticas adecuadas sobre privacidad y confidencialidad,

infraestructuras y una cultura de uso compartido de los datos. Si se consigue una buena implementación del Big data, dentro de poco, empezaremos a disfrutar de las espectaculares ventajas de la medicina del futuro.

Referencias

- MENASALVAS, E; GONZALO, C; RODRÍGUEZ-GONZÁLEZ, A. Big Data en Salud: Retos y oportunidades. Revista Economía Industrial Universidad Politécnica de Madrid, 2013,405: p. 87-97.
- Instituto de ingeniería del conocimiento. Big Data en medicina: aplicaciones útiles [En línea].2016 [Consultado el: 4 de abril de 2018] p. Disponible en: <http://www.iic.uam.es/lasalud/big-data-en-medicina-aplicaciones-utiles/>
- ARRIBAS VALIENTE, S. De la medicina traslacional a la medicina de precisión: el nuevo paradigma. Revista de la sociedad española de informática y salud, 2017, 124 p. 1-70.
- ALBIN, O; BARBA, G; DOMÍNGUEZ, E. Impacto de Big Data en la mejora de la calidad y en la imagen ética de las instituciones de salud Argentinas. facultad de administración y negocios, Universidad argentina de la empresa, Buenos Aires, 2014.
- FACUNDO MALVICINO. Big data aplicada al sector de salud en argentina. Definiendo una agenda de trabajo. CIECTI, 2016.
- MENDEZ GARCIA, M; RODA FAURA, M. Aproximación práctica al uso de las tecnologías Big datas en el sector de la salud. Comunicaciones, 2017, 27: 152-159.
- NOVAL GARCÍA, Y; ESPINOSA ACEREDA, K. Estudio comparativo de plataformas y herramientas de tratamiento de sistemas Big Data en Telemedicina. Ingeniería de Comunicaciones, Universidad del País Vasco, Bilbao, 2017.J
- MANUEL DE BUENAGA, D; MANUEL J.MAÑA, J; L. BORRAJO, L. IPHealth: Plataforma inteligente basada en open, linked y big data para la toma de decisiones y aprendizaje en el ámbito de la salud. Procesamiento del Lenguaje Natural, 2015, 55, p. 161.164.
- MERINO COSGAYA, D. Soluciones de privacidad y seguridad para diferentes escenarios de Big Data en medicina. Ingeniería de tecnologías específicas de telecomunicación, Universidad de Valladolid, Valladolid, 2016.
- GLORIA, P. Peligros del uso de los big data en la investigación en salud pública y en epidemiología [En línea]. 2016 vol.30 [consultado el: 3 de abril de 2018]. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112016000100014
- VICENTE PARDO, J.La Incapacidad laboral como indicador de gestión sanitaria [En línea]. 2015 vol.61 [consultado el: 3 de abril de 2018]. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0465-546X2015000200007
- GARCIA BARBOSA, J.La medician del futuro pasa por Big Data [En línea]. 2014 [consultado el: 3 de abril de 2018]. Disponible en: <https://aunclidelastic.blogthinkbig.com/la-medicina-del-futuro-pasa-por-big-data/>
- DE BURGOS, R.«Big data» y salud: preparados, listos... ¿ya? [En línea]. 2017 [consultado el: 3 de abril de 2018]. Disponible en: <http://elfarmaceutico.es/index.php/profesion/item/7911-big-data-y-salud-preparados-listos-ya#.WwChN0iFNPZ>
- Mejía. (2016). Crecimiento del Volumen de Datos en el Mundo. [Figura]. Recuperado de <https://es.slideshare.net/JuanCMejiaLlano/amosm2016-conferencia-social-big-data-por-juan-carlos-meja-llano>.