

# MODELOS DE TECNOLOGÍAS DEL BIG DATA ANALYTICS Y SU APLICACIÓN EN SALUD

***Gustavo Verduzco Reyes***

Universidad Popular Autónoma del Estado de Puebla  
*gvrmx@hotmail.com*

***Ernesto Bautista Thompson***

Universidad Popular Autónoma del Estado de Puebla  
*eb\_thompson@yahoo.com*

***Jorge A. Ruiz Vanoye***

Universidad Autónoma del Estado de Hidalgo  
*jorge@ruizvanoye.com*

***Alejandro Fuentes Penna***

Universidad Autónoma del Estado de Hidalgo  
*alexfp10@hotmail.com*

## **Resumen**

Big Data Analytics se refiere al almacenamiento, administración y análisis de grandes volúmenes de datos a través de métodos estadísticos o científicos para descubrir relaciones entre los datos. Recientemente se ha aplicado al campo de la salud, pero, ¿Qué tecnologías de big data aplicar a campos específicos en salud? El presente trabajo identifica áreas de oportunidad en salud que se benefician del big data Analytics, como la medicina personalizada, registros de salud, estancias y readmisiones de pacientes y biomedicina. También, se presenta una propuesta de modelo que conjunta plataformas tecnológicas (Hadoop, Mahout, Spark), algoritmos (Filtrado colaborativo, árboles de decisión, clustering) y campos de salud. Como ejemplo, se aplica el modelo propuesto en el caso del monitoreo a distancia de la salud de un paciente con problemas del corazón, como una base para su implementación real en un trabajo futuro.

**Palabras Claves:** Big data, modelo tecnológico, salud.

## **Abstract**

*Big Data Analytics refers to the storage, management and analysis of large volumes of data through statistical or scientific methods to discover relationships between data. Recently it has been applied to the field of healthcare, but, What big data technologies apply to specific fields in healthcare? This paper identifies areas of opportunity in healthcare that benefit from the big data analytics, such as personalized medicine, health records, stays and readmissions of patients and biomedicine. Also, a model proposal is presented that combines technological platforms (Hadoop, Mahout, Spark), algorithms (Collaborative filtering, decision trees, clustering) and health fields. An application of the model in relation to the remote monitoring of the patient's health with heart problems illustrates its use, as a basis for its implementation in future work.*

**Keywords:** *Big data, healthcare, technological model.*

## **1. Introducción**

Francis Diebold en el año 2003 fue el primero en utilizar el término de *Big Data* para explicar el fenómeno del crecimiento de datos [Diebold, 2012]. Big data se refiere a grandes conjuntos de datos que no tienen un solo formato, sino que contienen datos estructurados y no estructurados. A manera de ejemplo, tenemos los historiales del Internet, correos electrónicos, documentos de textos, transacciones comerciales, anuncios comerciales, análisis médicos, páginas web, etc. En [Jiang, 2015] se presentan cinco características que distinguen al *Big Data*: Veracidad (se refiere a que tan confiables y relevantes son los datos), Velocidad (la rapidez con la que se recolectan o generan los datos), Valor (se relaciona con la utilidad de los datos), Variedad (abarca distintos tipos, contenidos y formatos de datos), Volumen (se refiere a la gran cantidad de datos).

Tradicionalmente, *Big Data Analytics* se ha aplicado a campos como la toma de decisiones en las empresas [McAfee, 2012], análisis de grandes flujos de datos en tiempo real como twits, Facebook, ventas, etc. [Pigni, 2016], análisis de grandes volúmenes de datos en la investigación de mercados de consumo [Bosch, 2016], prevención de crímenes sexuales [Jin-ho, 2016], entre otros.

Recientemente los investigadores han mostrado interés en *big data analytics* aplicado al campo de la salud, de hecho [Feldman, 2012] señala que en 2012 se generaron a nivel mundial 500 petabytes de información de salud, esta información requiere ser analizada e interpretada para su valoración. Varios trabajos han identificado la aplicación de *big data* a la salud, en su análisis [Priyanka, 2014] identifica tres áreas específicas de salud como el soporte a decisiones clínicas, cuidados personalizados y operaciones clínicas. [Groves, 2013] encontraron como principales áreas de oportunidad en salud los datos clínicos, comportamiento del paciente y datos de farmacéutica (ensayos clínicos). Por su parte, [Hermon, 2014] en su estudio encontró cinco categorías: manejo de costos, soporte a decisiones clínicas, información clínica, comportamiento del paciente e Información de soporte. Las investigaciones se han centrado principalmente en encontrar las áreas de aplicación de *big data* en salud, sin embargo, hace falta identificar cómo las tecnologías de *big data* impactan esas áreas.

El propósito de este trabajo es mostrar cómo las tecnología asociadas al *big data analytics* pueden aplicarse en el campo de la salud. Se realizó un análisis de la literatura científica en relación a los principales campos de la salud y se generó un modelo teórico que integra ambas partes y un ejemplo práctico de su aplicación.

## 2. Métodos

### Big Data Analytics y sus Plataformas Tecnológicas

La analítica de datos (*data analytics*) se refiere a todo método matemático o científico que permita obtener una nueva visión de los datos o descubrir patrones de esos datos [Dietrich, 2014]. *Big data* y *data analytics* se han empleado para describir la relación entre conjuntos de datos y técnicas de análisis de datos para aplicaciones que requieren manejo de terabytes o exabytes de datos y donde es indispensable la tecnología para almacenar, administrar, y analizar datos.

La arquitectura de *Big data analytics*, figura 1, destaca el proceso que se sigue para tratar un gran conjunto de datos. Primero los datos son adquiridos de varias fuentes, internos de una misma organización o externos de otras organizaciones;

luego son curados (preparados) en formato de archivos de Excel, XML, CSV, y otros formatos no estructurados como archivos de texto, correos electrónicos, páginas web, twits. Posteriormente se aplican una serie de herramientas tecnológicas para manipular esos datos y finalmente se aplica la analítica de datos, en forma de consultas, reportes, técnicas de minería de datos, técnicas estadísticas. Los resultados de la analítica son presentados de forma gráfica mediante distintos tipos de diagramas, árboles, diagramas de pastel, líneas, 3D, etc.

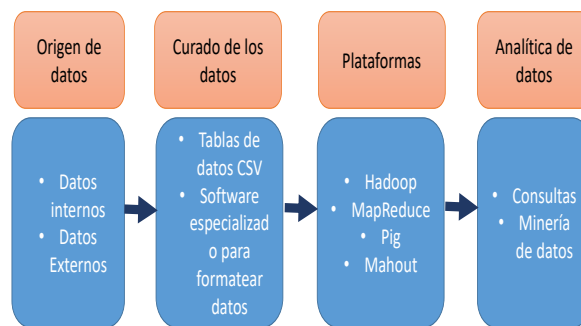


Figura 1 Arquitectura de *Big Data Analytics*.

Actualmente, con el poder de cómputo disponible se cuenta con una gran variedad de plataformas tecnológicas que sirven para tratar el *big data Analytics*, tabla 1, estas permiten almacenar y procesar flujos de datos de gran magnitud.

Toda vez que se cuenta con la información almacenada y administrada a través de alguna plataforma tecnológica, se analiza a través de algoritmos y técnicas, ver tabla 2, que nos permitan obtener patrones de comportamiento o información significativa. Generalmente estos algoritmos se enmarcan dentro de la minería de datos.

Por otro lado, la producción de información sobre enfermedades y su tratamiento no se detiene, de acuerdo a [Ivan, 2015] se estima que la producción de datos médicos se incrementará en un 4,300%, esto es, de 2.52 zettabytes en 2010 a 73.5 zettabytes en 2020. Recientemente se ha enfocado el *Big Data Analytics* al campo de la salud, los recursos de información disponible para pacientes, cada vez está creciendo más y no es posible tratarlos con las técnicas tradicionales de bases de datos relacionales.

Tabla 1 Plataformas de *Big Data Analytics*.

Plataforma	Descripción
Hadoop	Maneja grandes volúmenes de datos y distribuye la carga de procesamiento de datos en los <i>clusters</i> de servidores, está compuesto por dos módulos, MapReduce y DFS (Administrador de archivos distribuido).
PIG	Es un lenguaje de alto nivel para flujo de datos, paraleliza grandes volúmenes de datos. Produce secuencia de programas del tipo <i>Map-Reduce</i> que pueden ser interpretadas por Hadoop.
MAHOUT	Es una biblioteca escalable para realizar minería de datos y aprendizaje máquina. Sus algoritmos son compatibles con Apache Spark, H2O y Apache Flink. Soporta multiplicaciones vectoriales grandes, su entorno para experimentos es similar al lenguaje de programación R.
HBase	Es un administrador de bases de datos distribuidos, con manejo de datos estructurados a gran escala, las tablas pueden contener miles de millones de filas y millones de columnas. Contiene un API fácil de usar para el lado del cliente.
Spark	Es un motor de manejo de datos que usa Hadoop, cuenta con soporte para aplicaciones de aprendizaje máquina, procesamiento de flujo de datos y visualización. Es compatible con lenguajes como Python y lenguaje R. También combina SQL y analítica compleja.
Cassandra	Es un gestor de base de datos a gran escala para datos de misión crítica. Permite escalabilidad lineal y probada tolerancia a fallos. Hace réplicas en múltiples centros de datos.

Para este trabajo se realizó un análisis de diversos documentos de investigación científica tomados de bibliotecas digitales como EBSCOhost, Dialnet, Web Of Science y Google Académico, las palabras claves que se emplearon fueron “*big data healthcare*”, “*big data application to health*”, “*big data analytics*”. Se halló que varios son los campos de la salud que se pueden beneficiar del *big data Analytics* como se explica a continuación:

- *Medicina personalizada*: Aunque actualmente se han desarrollado muchos medicamentos para distintas enfermedades, se requiere tener diagnósticos óptimos que permitan determinar que enfermedad tiene un paciente y proporcionar un tratamiento adecuado. Por tanto, es necesario un entendimiento profundo de una enfermedad basado en su comportamiento genético [Hamburg, 2010].

Tabla 2 Algoritmos para el análisis de datos.

Propósito	Algoritmo de análisis
Filtrado Colaborativo	<p><b>Filtrado colaborativo basado en el usuario.</b>- Permiten generar un perfil de usuario en base a su comportamiento digital y luego encontrar otros usuarios con gustos similares.</p> <p><b>Filtrado colaborativo basado en Elemento.</b>- Permite que un usuario encuentre elementos similares que gustaron a otro usuario.</p>
Clasificación	<p><b>Naive Bayes.</b>- Es un clasificador estadístico, basado en el teorema de Bayes. Ayuda a predecir si una entidad pertenece a una clase. Puede manejar grandes cantidad de datos con rapidez y precisión.</p> <p><b>Bosque aleatorio.</b>- Clasificador para grandes volúmenes de datos, introduce la aleatoriedad en cada clasificador individual. Parte de lo que se conoce como árbol de decisión.</p> <p><b>Modelos ocultos de markov.</b>- Permite encontrar parámetros desconocidos (ocultos) a partir de parámetros observables. Se puede ver como un doble proceso estocástico.</p> <p><b>Perceptron multicapa.</b>- Se considera un algoritmo que realiza un aprendizaje por corrección de error. Permite un alto grado de conectividad en sus capas o neuronas. Puede usarse para predecir estancias hospitalarias de pacientes.</p>
Agrupamiento (clustering)	<p><b>K-Means clustering.</b>- Permite dividir un conjunto de datos en subconjuntos (<i>clusters</i>). Generalmente, se emplea la distancia Euclidiana para establecer la cercanía del <i>cluster</i> a un centroide y de esta manera identificar a qué <i>clúster</i> pertenece un elemento.</p> <p><b>Fuzzy k-Means.</b>- Los elementos de un conjunto de datos tienen un grado de pertenencia a un clúster. Los elementos no son dicotómicos al no estar sujetos a pertenecer de forma definitiva a un <i>clúster</i>.</p> <p><b>Spectral Clustering.</b>- Usa la matriz de similitud de datos para reducir dimensionalidad. Los datos son representados como un grafo, la unión de los vértices tiene un peso. A mayor peso, se interpreta como mejor similitud, menor peso, menor similitud.</p>

Analizar características de enfermedades y crear perfiles de las mismas, que empaten con los síntomas de un paciente es una tarea que requiere herramientas tecnológicas para analizar grandes conjuntos de datos heterogéneos.

- **Registros de salud:** Estos registros se refieren a los historiales de salud de pacientes, fechas de ingreso a hospitales, enfermedades adquiridas, padecimientos, tratamientos a los que se ha sometido, notas médicas, resultados clínicos, entre otros datos. Aunque los centros de atención de salud en el mundo no cuentan con un estándar único para los registros [Mao, 2017], es importante analizar la información de distintos hospitales

para comprender el comportamiento de cierta enfermedad y cómo se ha tratado.

- *Estancias y readmisión de pacientes:* Muchos pacientes son internados, tratados u operados en hospitales, sin embargo, luego de haber sido dados de alta, retornan a los pocos días, esto representa un costo para los pacientes, hospital y aseguradoras. En algunos casos, hasta se pueden generar demandas legales de parte de los familiares de un paciente por la falta de atención expedita [Buffa, 2007]. La tecnología de *big data Analytics* podrían proveer técnicas de análisis de datos para minimizar la reincidencia hospitalaria.
- *Biomedicina:* Gracias al secuenciado del genoma humano, ahora es posible analizar enfermedades para comprender su estructura a nivel celular y molecular, para generar fármacos más efectivos. El reto es integrar y analizar los repositorios de información disponibles con tecnología más rápida y más potente, para analizar e interpretar los resultados obtenidos [Costa, 2014].

Un análisis basado en las características de las tecnologías de *big data analytics* y las diversas áreas de salud han permitido generar una tabla, ver tabla 3, donde se resalta cuál es la tecnología de *big data* que puede aplicarse en salud. Las tecnologías de *big data analytics* como Hadoop, Mahout, Pig son robustas, al igual que las técnicas de análisis como Filtrado Colaborativo, *Clustering*, *Naive Bayes* y Árboles de Decisión. Desde luego, a medida que se integren más trabajos en este tipo de análisis, será más claro el impacto del *big data Analytics* a la salud.

### 3. Resultados

La tecnología para el almacenamiento y procesamiento de datos en su mayoría es de código abierto, lo mismo que los algoritmos para aplicar la analítica de datos en salud. Esto ha permitido que muchos investigadores contribuyan a encontrar patrones de comportamiento de enfermedades, así como posibles soluciones para su tratamiento.

Tabla 3 Tecnologías de *Big Data* aplicadas a Salud.

Autor	Campo de salud	Tecnología <i>Big Data</i> Aplicable
[Belle et al., 2015]	Procesamiento de imágenes, procesamiento de señales, información genómica.	Hadoop, <i>clustering</i>
[Chawla & Davis, 2013]	Medicina personalizada	Mahout, Filtrado Colaborativo
[Simpao, Ahumada, & Rehman, 2015]	Apoyo a decisiones clínicas y gestión de recursos	Hbase, árboles de decisión
[Luo, Wu, Gopukumar, & Zhao, 2016]	Análisis genómico	Mahout, clasificación Naive Bayes
[Murdoch & Detsky, 2013]	Registros de salud electrónico	Spark, Filtrado colaborativo
[Obenshain, 2004]	Identificar pacientes de alto riesgo y control de infecciones	Mahout, Filtrado basado en el usuario
[Raghupathi & Raghupathi, 2014]	Monitoreo de dispositivos o monitoreo remoto	Mahout, Naive Bayes
[Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014]	Estancias y Readmisión de pacientes	Hadoop, Naive Bayes
[Schneeweiss, 2014]	Análisis de estudios clínicos	PIG, Árboles de decisión
[Costa, 2014]	Biomedicina	Spark, Spectral clustering

En la industria hay dos modelos importantes para el análisis de datos, en [Balkan, 2010] se hace referencia al modelo SEMMA por sus siglas en inglés (Sample, Explore, Modify, Model, Assess) el cual consta de cinco fases Muestra, Explorar, Modificar, Modelar, Evaluación. El segundo, es el modelo CRISP-DM (CRoss Industry Standard Process for Data Mining), para [Hiltbrand, 2013] consta de las siguientes fases: comprender los datos, preparación de los datos, modelado, evaluación, producción y entender el negocio. Basado en los modelos anteriores, a continuación se presenta una propuesta de modelo de *big data Analytics* y salud, figura 2, que considera tres grandes pilares: plataformas tecnológicas, algoritmos y campos de salud.

*Plataformas tecnológicas:* En el presente texto se han abordado aquellas que son de carácter libre (Hadoop, Pig, Mahout), sin embargo, empresas como Amazon, Google y Microsoft, principales gigantes del software a nivel mundial, prestan



servicios en la nube que son configurables según la necesidad de almacenamiento de datos y uso de recursos de procesamiento.

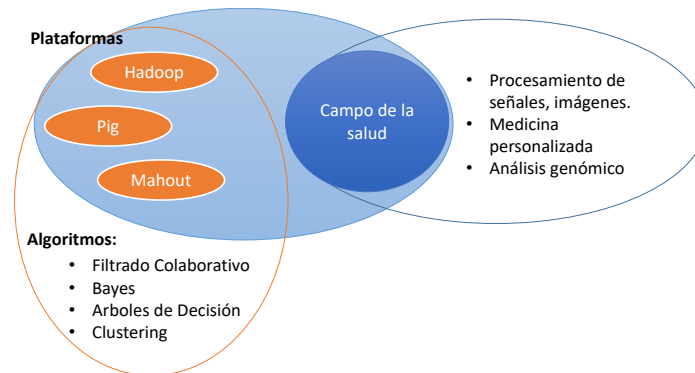


Figura 2 Modelo de *Big Data Analytics* y Salud.

**Algoritmos:** Estos han demostrado ser efectivos al tratar grandes cantidades de datos, como detección de correos spam, análisis de sentimientos en redes sociales, clasificación de documentos, marketing, etc. Su aplicación en el campo de la salud también está dando buenos resultados, sobre todo en la detección y prevención de enfermedades. [Lazer, 2014] mencionan que la empresa Google ha utilizado la algorítmica para comprender el comportamiento de la influenza.

**Campos de la salud:** Hemos descrito algunos de los campos más representativos que pueden beneficiarse del uso del *big data Analytics*. Sin embargo, otros como el tratamiento de señales en tiempo real generados por dispositivos electrónicos médicos, análisis de imágenes (rayos X, tomografía, etc.) y análisis epidemiológico también pueden beneficiarse de esta tecnología.

### Ejemplo de Aplicación del Modelo Propuesto

Un paciente que egresa de un hospital requiere en muchas ocasiones que se le dé seguimiento constante de su salud, máxime si le ha sido implantado uno o más dispositivos electrónicos para monitorear su condición actual de salud. El monitoreo a distancia puede ayudar a reducir la asistencia innecesaria a consultas hospitalarias, también puede prevenir un posible descontrol en el paciente y alertar de esta condición [Verma, 2015]. Con objeto de aplicar el modelo propuesto a un caso específico (figura 3) de monitoreo remoto para medicina personalizada. En

nuestro ejemplo, un hospital cuenta con una plataforma de monitoreo remoto para una población amplia de pacientes, esto sugiere una gran cantidad de datos generados por cada uno de ellos. Ahora pensemos en un paciente que sufre problemas cardíacos y que le han sido implantado uno o más de los siguientes dispositivos para realizarle un monitoreo a distancia:

- *Marcapasos*.- Es un dispositivo electrónico alimentado por baterías que envía señales eléctricas al corazón cuando éste late muy rápido o muy lento para regularizarlo a un ritmo normal.

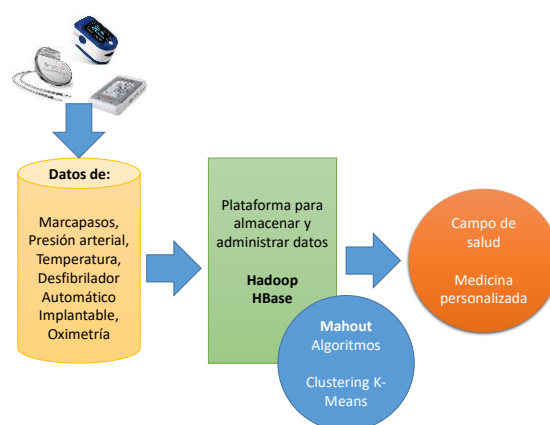


Figura 3 Ejemplo del Modelo propuesto.

- *Dispositivo de Presión arterial*.- También denominado de Presión Sanguínea, se emplea para medir la presión que ejerce la sangre en las arterias. Algunos envían los resultados vía wifi a un servidor de recepción de datos.
- *Dispositivo de control de temperatura*.- Se emplea para medir la temperatura corporal de un individuo. Algunos dispositivos sofisticados se conectan vía Bluetooth a un teléfono celular para luego enviar los datos a un servidor de datos cada cierto tiempo, según sea programado.
- *Desfibrilador Automático Implantable*.- Se emplea para detener y revertir arritmias cardíacas a través de un choque eléctrico. Algunos dispositivos logran hacer un censo de datos de 30, 60, 90 días.
- *Dispositivo de oximetría*.- Se emplean para medir los niveles de oxígeno en la sangre, detectan la condición del corazón y pulmones.

Debido a que los datos generados por los dispositivos se realizan por horas, días, semanas o meses requieren almacenarse para su análisis, en vista que son datos heterogéneos o requieren gran capacidad de almacenamiento. Hadoop es una tecnología de manejo de archivos a escala y Hbase se emplea para manipular tablas de datos de millones de registros, tal como se observa en la figura 4 los datos generados por los dispositivos se crean a través de instancias de almacenamiento de datos llamadas RegionServer las cuales no son otra cosa que tablas de datos, se pueden crear de forma dinámica según se requieran, éstas son almacenadas de forma lógica en un servidor Hadoop bajo el formato del sistema de archivos HDFS. En cuanto al crecimiento de los *cluster* de almacenamiento este también es dinámico y es organizado por un *cluster* líder de Hadoop.

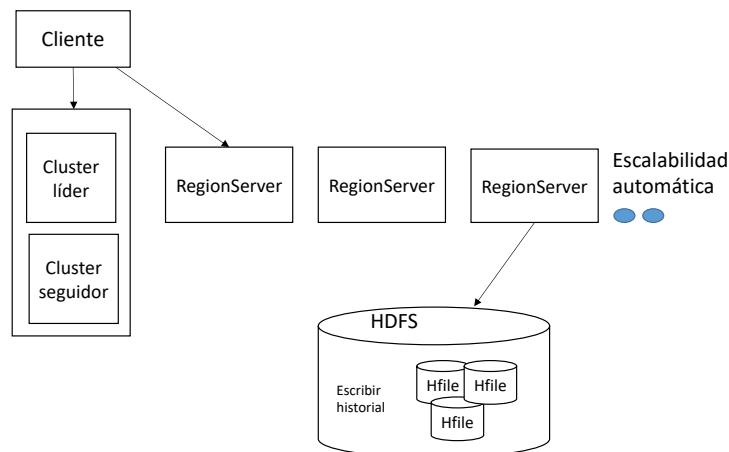


Figura 4 Almacenamiento y manejo de datos.

El análisis de los datos almacenados es una tarea que puede ser llevada a cabo en Mahout, librería de aprendizaje máquina compatible con Hadoop y que se maneja una gran cantidad de algoritmos. Para este caso tomaremos el algoritmo de clasificación K-Means, supongamos que se quiere revisar la presión arterial del paciente para detectar un posible ataque al corazón, de acuerdo a la figura 5 se analizan los datos arrojados por el sensor que toma la presión, se revisan los parámetros de la presión arterial sistólica y la presión arterial diastólica. El algoritmo K-Means permite tomar estos dos valores y por medio de repeticiones sucesivas generaría grupos en base a un valor de referencia denominado

centroide, cuanto más cerca estén los grupos de un centroide significan que pertenecen a ese grupo.

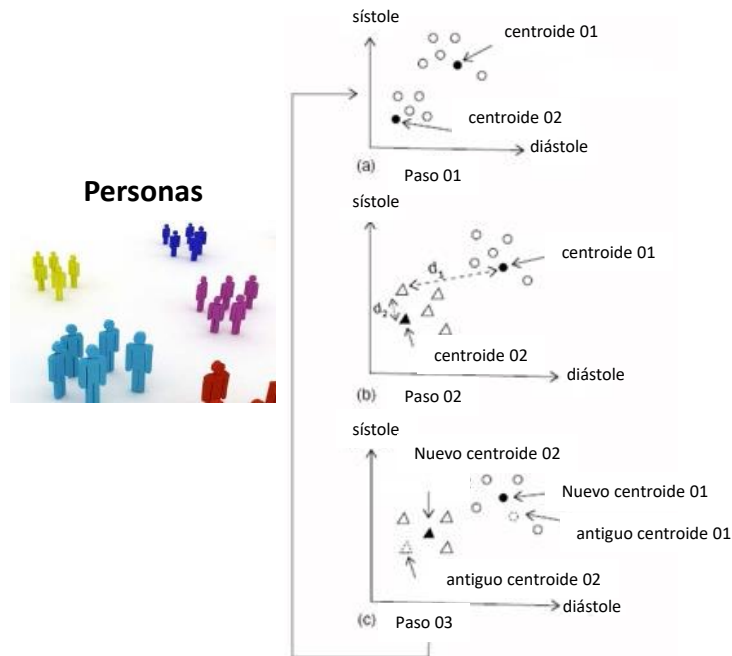


Figura 5 Algoritmo K-Means para clasificación

Una vez que se ha aplicado la analítica de los datos, se pueden interpretar los resultados por un especialista en salud y generar medicina personalizada para el paciente, crear un tratamiento acorde a su condición actual de salud. Por tanto, la observación permanente a distancia de los dispositivos permite evitar que el paciente se mueva de forma innecesaria de su casa al hospital para recibir tratamiento médico sino es requerido. También, permite ver el progreso de recuperación de un paciente de forma más cercana y avisarle de algún posible peligro.

#### 4. Discusión

El presente trabajo ha identificado áreas de oportunidad donde se pueden aplicar las tecnologías de *big data Analytics* a la salud, también presenta un modelo de arquitectura de *big data* que combina el campo de la salud con tecnologías *big data* específicas de almacenamiento y análisis de datos.

Por otro lado, la aplicación de la propuesta de modelo al caso de estudio de monitoreo a distancia con un paciente de problemas de corazón, guarda relación con otros estudios, tal como señala [Srinivasan, 2013] donde aplicaron una serie de modelos predictivos sobre historiales médicos para detectar reclamos por admisiones hospitalarias. También [O'Driscoll, 2013] presenta el uso de Apache Hadoop aplicado en el campo de la medicina genómica, en particular los algoritmos para el análisis de secuencias genómicas y análisis de RNA y [Andreu, 2015] resaltan que *big data Analytics* se puede aplicar a campos como la proteómica, genómica, farmacología.

El análisis empleado en este trabajo, permite obtener información útil de las características de las tecnologías de *big data analytics*, así como de los campos de la salud. Evidentemente, es necesario explorar otros algoritmos de análisis que permitan beneficiar a otros campos de la salud como el tratamiento de imágenes o de audio.

Por otro lado, es necesario investigar cuál es el impacto de estas tecnologías de *big data Analytics* en el campo de la salud. También, se requiere determinar si la comunidad médica considera útiles la aplicación de estas tecnologías y si los conducen a nuevos esquemas de tratamiento de enfermedades.

## 5. Conclusiones

Este estudio ha identificado cuáles son las tecnologías de *big data Analytics* que pueden aplicarse en áreas específicas de salud. El modelo de arquitectura presentado da cuenta que la aplicación de la tecnología de *big data* apoya en las tareas de detección, análisis y toma de decisiones para un combate más responsivo de diversas enfermedades. Lo anterior tiene implicaciones significativas, ya que, al combinar los datos de salud almacenados, con los generados en tiempo real ayudará a comprender mejor el comportamiento de enfermedades, por ejemplo, la influenza y otras enfermedades pandémicas.

Esta investigación, no considera todos los algoritmos que pueden aplicarse en áreas específicas de salud sino los más utilizados para la práctica de *big data Analytics*, por tanto, recomendamos más investigación a este respecto, lo que

puede incluir una revisión de las mejores prácticas de la algorítmica con enfoque específico de salud.

A futuro se busca colaborar con el sector salud y tener acceso a bases de datos médicas para la puesta en marcha y evaluación operativa del modelo propuesto, utilizando en su implementación herramientas libres de *big data*. Asimismo, medir y comparar la eficacia de los algoritmos de análisis de información utilizados en el modelo a implementar. Finalmente, evaluar, por ejemplo, con un escalamiento Likert, la utilidad del modelo operativo para los usuarios de la comunidad médica.

## 6. Bibliografía y Referencias

- [1] Andreu Perez, J., Poon, C. C., Merrifield, R. D., Wong, S. T., & Yang, G.-Z., Big data for health. *IEEE Journal of Biomedical and Health Informatics*, 19(4), pp.1193–1208, 2015.
- [2] Balkan, S., & Goul, M., Advances in Predictive Modeling: How In-Database Analytics Will Evolve to Change the Game. *BUSINESS INTELLIGENCE JOURNAL*, 15(2), 2010: [https://www.researchgate.net/profile/Sule\\_Balkan/publication/264905640\\_Advances\\_in\\_Predictive\\_Modeling\\_How\\_InDatabases\\_Analytics\\_Will\\_Evolve\\_to\\_Change\\_the\\_Game/links/53f5c3870cf22be01c3faa29.pdf](https://www.researchgate.net/profile/Sule_Balkan/publication/264905640_Advances_in_Predictive_Modeling_How_InDatabases_Analytics_Will_Evolve_to_Change_the_Game/links/53f5c3870cf22be01c3faa29.pdf).
- [3] Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G., Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, 33(7), pp. 1123–1131, 2014.
- [4] Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K., Big Data Analytics in Healthcare. *BioMed Research International*, 2015, pp.1–16, 2015, <https://doi.org/10.1155/2015/370194>.
- [5] Bosch, V. (2016). Big Data in Market Research: Why More Data Does Not Automatically Mean Better Information. *GfK-Marketing Intelligence Review*, 8(2), pp. 56–63, 2016, <https://doi.org/10.1515/gfkmir-2016-0017>.
- [6] Chawla, N. V., & Davis, D. A., Bringing big data to personalized healthcare: a patient-centered framework. *Journal of General Internal Medicine*, 28(3), pp. 660–665, 2013.

- [7] Buffa, M., Petti, M., & Perasso, O., Readmisión hospitalaria en un centro ambulatorio quirúrgico. *Rev. Argent. Anestesiol*, 65(1), pp. 30–40, 2007.
- [8] Costa, F. F., Big data in biomedicine. *Drug Discovery Today*, 19(4), pp. 433–440, 2014.
- [9] Diebold, F. X., On the Origin (s) and Development of the Term 'Big, 2012 Data': [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2152421](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2152421).
- [10] Dietrich, B. L., Plachy, E. C., & Norton, M. F., *Analytics across the enterprise: How IBM realizes business value from big data and analytics*, 2014 IBM Press: <https://books.google.com/books?hl=es&lr=&id=IGOVAAwAAQBAJ&oi=fnd&pg=PR13&dq=Analytics+across+the+enterprise:+how+ibm+realizes+business+value+from+big+data+and+analytics&ots=cKqJPCO-7N&sig=1A4GHykiBRN8c5bfx7-nLCqUMyg>.
- [11] Feldman, B., Martin, E. M., & Skotnes, T., Big Data in Healthcare Hype and Hope. *Dr. Bonnie*, 360, 2012: [http://www.ghdonline.org/uploads/big-data-in-healthcare\\_B\\_Kaplan\\_2012.pdf](http://www.ghdonline.org/uploads/big-data-in-healthcare_B_Kaplan_2012.pdf).
- [12] Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S., The “big data” revolution in US healthcare. McKinsey & Company Web site, 2013.
- [13] Hamburg, M. A., & Collins, F. S., The path to personalized medicine. *New England Journal of Medicine*, 363(4), pp. 301–304, 2010.
- [14] Hermon, R., & Williams, P. A., Big data in healthcare: What is it used for?, 2014: <http://ro.ecu.edu.au/aeis/22/>.
- [15] Hiltbrand, T., Behavior-Based Budget Management Using Predictive Analytics. *The Business Intelligence Journal*, 18(INL/JOU-12-26713), 2013: <http://www.osti.gov/scitech/biblio/1072389>.
- [16] Ivan, M., & Velicanu, M., Healthcare industry improvement with business intelligence. *Informatica Economica*, 19(2), 81, 2015.
- [17] Jiang, F., & Leung, C., A Data Analytic Algorithm for Managing, Querying, and Processing Uncertain Big Data in Cloud Environments. *Algorithms*, 8(4), pp. 1175–1194, 2015: <https://doi.org/10.3390/a8041175>.
- [18] McAfee, A., & Brynjolfsson, E., Big data: the management revolution. *Harvard Business Review*, 90(10), 60, 2012.

- [19] Jin-ho Jeon, & Seung-Ryul Jeong, Designing a Crime-Prevention System by Converging Big Data and IoT. *Journal of Korean Society for Internet Information*, 17(3), pp. 115–128: <https://doi.org/10.7472/jksii.2016.17.3.115>.
- [20] Lazer, D., Kennedy, R., King, G., & Vespignani, A., The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176), pp. 1203–1205, 2014: <https://doi.org/10.1126/science.1248506>.
- [21] Luo, J., Wu, M., Gopukumar, D., & Zhao, Y., Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomedical Informatics Insights*, 8, 1, 2016.
- [22] Mao, H., & Sun, Y., A Way to Understand Inpatients Based on the Electronic Medical Records in the Big Data Environment. *International Journal of Telemedicine & Applications*, 1–9: <https://doi.org/10.1155/2017/9185686>.
- [23] Murdoch, T. B., & Detsky, A. S., The inevitable application of big data to health care. *JAMA*, 309(13), 1351–1352, 2013: <https://doi.org/10.1001/jama.2013.393>.
- [24] Obenshain, M. K., Application of data mining techniques to healthcare data. *Infection Control & Hospital Epidemiology*, 25(08), pp. 690–695, 2014.
- [25] O'Driscoll, A., Daugelaite, J., & Sleator, R. D., Big data, Hadoop and cloud computing in genomics. *Journal of Biomedical Informatics*, 46(5), pp. 774–781, 2013.
- [26] Pigni, F., Piccoli, G., & Watson, R., Digital Data Streams: CREATING VALUE FROM THE REAL-TIME FLOW OF BIG DATA. *California Management Review*, 58(3), pp. 5–25, 2016.
- [27] Priyanka, K., & Kulennavar, N., A survey on big data analytics in health care. *International Journal of Computer Science and Information Technologies*, 5(4), 5865–8, 2014.
- [28] Raghupathi, W., & Raghupathi, V., Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 1, 2014.
- [29] Simpao, A. F., Ahumada, L. M., & Rehman, M. A. (2015). Big data and visual analytics in anaesthesia and health care. *British Journal of Anaesthesia*, aeu552.



- [30] Schneeweiss, S., Learning from big health care data. *New England Journal of Medicine*, 370(23), pp. 2161–2163, 2014.
- [31] Srinivasan, U., & Arunasalam, B., Leveraging big data analytics to reduce healthcare costs. *IT Professional*, 15(6), pp. 21–28, 2013.
- [32] Verma, A., & Yu, C.-M., HRS Expert Consensus Statement on remote interrogation and monitoring for cardiovascular implantable electronic devices. *Heart Rhythm*, 12(7), e70, 2015.