



Revista Venezolana de Gerencia
ISSN: 1315-9984
rvgluz@gmail.com
Universidad del Zulia
Venezuela

Big Data y las implicaciones en la cuarta revolución industrial - Retos, oportunidades y tendencias futuras

Araque González, Gustavo Andrés; Gómez Vásquez, Mauricio; Vélez Uribe, Juan Pablo; Suárez Hernández, Albeiro Hernán

Big Data y las implicaciones en la cuarta revolución industrial - Retos, oportunidades y tendencias futuras

Revista Venezolana de Gerencia, vol. 26, núm. 93, 2021

Universidad del Zulia, Venezuela

Disponible en: <https://www.redalyc.org/articulo.oa?id=29066223003>

Big Data y las implicaciones en la cuarta revolución industrial - Retos, oportunidades y tendencias futuras

Big Data and the implications of the fourth industrial revolution - Challenges, opportunities and future trends: A review

Araque González, Gustavo Andrés
Universitaria Politécnico Grancolombiano, Colombia
garaque@poligran.edu.co

Redalyc: <https://www.redalyc.org/articulo.oa?id=29066223003>

Gómez Vásquez, Mauricio
Universitaria Politécnico Grancolombiano, Colombia
mgomezva@poligran.edu.co

Vélez Uribe, Juan Pablo
Universitaria Politécnico Grancolombiano, Colombia
jpvelezu@poligran.edu.co

Suárez Hernández, Albeiro Hernán
Universitaria Politécnico Grancolombiano, Colombia
asuarez@poligran.edu.co

RESUMEN:

La cuarta revolución industrial se presenta, en la actualidad, como un proceso de transformación y evolución productiva. El comportamiento globalizado de los mercados, las nuevas tendencias tecnológicas y el auge de innovadoras metodologías, han transformado la industria en un espacio de interacción interdisciplinar para la toma de decisiones organizacionales. Uno de estos espacios es la generación, obtención y análisis de los datos obtenidos de acuerdo con la relación hombre-máquina, también conocido en la actualidad como el Big Data. El presente artículo de investigación descriptiva tiene por objetivo exponer escenarios de aplicación, control y análisis de técnicas, tecnologías y metodologías asociados al Big Data en tres sectores principales: salud, financiero y transporte y logística. Los resultados obtenidos permiten evidenciar el impacto y relevancia con relación al mundo Big Data y sus aplicaciones en los sectores industriales de estudio como eje de reflexión y repercusiones profesionales. Se concluye que la integración de hardware y software en el campo del Big Data se hace indispensable en pro de la mejora de la calidad de servicio de teleasistencia en los pacientes, así como en los servicios logísticos y financieros.

PALABRAS CLAVE: industria 4, 0; macro datos; internet de las cosas; minería de datos.

ABSTRACT:

The fourth industrial revolution is currently presented as a process of transformation and productive evolution. The globalized behavior of markets, new technological trends and the rise of innovative methodologies have transformed the industry into an interdisciplinary interaction space for organizational decision making. One of these spaces is the generation, collection and analysis of the data obtained according to the man-machine relationship, also known today as Big Data. This descriptive research article presents some scenarios of application, control and analysis of techniques, technologies and methodologies associated with Big Data in three main sectors: health, financial and transportation and logistics. The results obtained show the impact and relevance in relation to the Big Data world and its applications in the industrial sectors of study as an axis of reflection and professional repercussions.

KEYWORDS: industry 4, 0; big data; internet of things; data mining.

1. INTRODUCCIÓN

Nuevos e innovadores escenarios de desarrollo y evolución industrial surgen como resultado de la transformación y evolución de los sistemas de producción y manufactura. Elementos como la inteligencia

artificial, el Internet de las Cosas (IoT), la manufactura aditiva (impresión 3D) y la era de digitalización de datos (Big Data) se presentan como fuentes y oportunidades de desarrollo investigativos en pro y mejora de la calidad de productos y servicios.

La presente investigación se enfoca en un análisis descriptivo de las principales tendencias y comportamientos de los datos informativos (Big Data), su análisis e interpretación (data mining) y sistemas de mejoramiento continuo y aprendizaje (Machine Learning) en las diversas aplicaciones para los sectores de salud, financiero y logística y transporte (Din et al, 2019) (Huber et al, 2019)

En relación con (Abugabah et al, 2020), uno de los desafíos en relación a la implementación y estructuración del Big Data en los sectores industriales de estudio se enfoca en la naturaleza no estructurada de las informaciones obtenidas de los procesos e interacción hombre-máquina: Fuentes de información de entrada, como por ejemplo sensores en dispositivos, sistemas de entradas por Radiofrecuencia (RFID), mecanismo de búsqueda en internet, procesamiento y generación de datos bancarios, médicos, gubernamentales, académicos, entre otros, se presentan como datos relevantes con diversos tipos de características (numérico, textual, gráfico, video, etc.), sin un control de salida definido y como información no estructurada. Y es esta última característica de datos la que ha despertado especial interés en la comunidad científica en el desarrollo de estudios y búsqueda de patrones estándar del comportamiento de la información de salida e interpretación y análisis de los mismos, (Shakhovska et al, 2019).

En el mundo de la Industria 4.0, la correcta utilización de los factores de eficiencia, eficacia y exactitud en relación a la información de salida del Big Data, son esenciales para la toma de decisiones empresarial (Frank et al, 2019) (Vieira et al, 2020).

La sincronía e integralidad de los datos obtenidos y su óptimo tratamiento, análisis y desarrollo demandan la existencia de profesionales especializados (arquitectos de datos, médicos genetistas informáticos, bioinformáticos, etc.), generación de trabajos multidisciplinarios orientados al estudio y desarrollo de datos, acompañados de inversión en infraestructura y tecnologías asociadas.

La presente investigación, expone un análisis descriptivo y aplicaciones del Big Data en tres sectores industriales: Salud, financiero y logística y transporte. Se presentan metodologías de desarrollo y tendencias tecnológicas enfocadas en cada una de las industrias, oportunidades y mejoras desarrolladas. La presentación de nuevas tendencias y estructuración de datos informativos desde cada perspectiva de análisis generan la apertura a nuevas e innovadoras oportunidades para los profesionales del sector industrial en el desarrollo de nuevas técnicas, metodologías y herramientas de aprendizaje, generando espacios de reflexión, generación de conclusiones y presentación de los principales exponentes bibliográficos.

2. Manufactura de servicios enfocados en la industria 4.0: Big data en el sector de salud. Implicaciones electrónicas – inteligencia clínica

Una de los elementos esenciales en los grandes entornos de desarrollo de la industria 4.0 se enfoca en las aplicaciones e innovaciones centralizadas en la administración y gestión de la información en bases de datos, conocido como Big Data. De acuerdo con (Zhang, 2018), el Big Data puede ser concebido como una tecnología que busca, como objetivo principal, el tratamiento de diversos tipos de información (estructurada, semiestructurada y desestructurada) de carácter dinámico. Patil & Seshadri (2014), argumentan que este tipo de informaciones pueden ser obtenidas de varias fuentes de entrada, como sensores, instrumentos o herramientas y de procesos transaccionales llevados a cabo en las computadoras. Una vez procesadas, estas se actualizan de forma constante, cuyo objetivo principal es la presentación de datos reales, actualizados y precisos, para su posterior estudio y procesos de investigación.

A esta etapa final de análisis de la información es lo que se conoce hoy en día como la minería de datos (*data mining*).

CUADRO 1
Fuentes y tipos de datos generados a partir de Big Data

Cuadro 1 Fuentes y tipos de datos generados a partir de Big Data	
Recursos de entrada de información	Tipo de información Generada
Correo electrónico, SMS, mensaje instantáneo, YouTube, WhatsApp, Web	Video textual, gráfico y audio
Instrumentos médicos electrónicos, científicos datos experimentales y observacionales	Númérico (es decir, temperatura, presión, etc.) e imágenes de diagnóstico (es decir, tomografía computarizada, ECG, etc.)
Sensores ambientales	Númérico, textual, gráfico, audio-video
Transacciones financieras	Textual y Numérico
Base de datos tradicional y Datawarehouse	Númérico y textual
Satélite	Númérico y gráfico

Fuente: Giacalone & Scippacercola (2016).

Giacalone & Scippacercola (2016).

En el mundo del Big data, de acuerdo con Perera et al, (2014) existen diversas categorías y presentación de la información para su correcto uso en la toma de decisiones. Una primera categoría de la información puede ser concebida como la relación e integración que existe entre los recursos de entrada y los productos obtenidos a partir de estas (también conocidas como fuentes de información generada). Lo anterior puede ser observado en la Cuadro 1.

Una de las aplicaciones principales de Inteligencia clínica y expuestas por (Übeyli, 2007) plantea el desarrollo integrado de sistemas de información de cinco grandes grupos de modelos cuantitativos: redes neuronales combinadas (Combined Neural Networks - CNNs), mezcla de expertos (Mixture of Experts - ME), redes neuronales probabilísticas (Probabilistic Neural Networks- PNN), redes neuronales recurrentes (Recurrent Neural Networks -RNN) y máquinas de vectores de soporte (Support Vector Machines -SVM) y redes neuronales de perceptrón multicapa (Multilayer Perceptron Neural Networks- MLPNNs), siendo este último el sistema cuantitativo utilizado como herramienta de estudio en la presente investigación.

CUADRO 2
Características de Volumen, Velocidad y Variedad de Big Data

Cuadro 2 Características de Volumen, Velocidad y Variedad de Big Data

Característica	Tamaño, tiempo y tipo de información
Volumen	Tamaño: TB (terabyte = 1024^4 byte) - PB (petabyte = 1024^5 byte) - EB (exabyte = 1024^6 byte) - ZB (zettabyte = 1024^7 byte) - YB (yottabyte = 1024^8 byte)
Velocidad	Tiempo: resultados en tiempo real: la adquisición rápida y el acceso a los datos es esencial, especialmente para 'datos en vivo' que deben procesarse diariamente.
Variedad	Tipo de datos: Estructurado - Semi-estructurado - No estructurado (cualitativo)

Fuente: Giacalone & Scippacercola (2016).

Giacalone & Scippacercola (2016).

Como fase de entrada, se establecen los datos fuentes, también conocidos como suministro de información, de característica no independiente y por el contrario con un alto grado de correlación entre ellos, en donde modificaciones en los resultados pueden generar cambios en las configuraciones en el sistema. Este tipo de información es ilustrada en el cuadro 2.

Lo anteriormente mencionado, puede llevarse a cabo a través de la creación de algoritmos de tratamiento de información enfocados en el área de servicio de estudio, y un primer paso para entender el sistema es la generación del concepto de Inteligencia Clínica. Un ejemplo de Inteligencia clínica aplicada al sector de la salud es ilustrado en el diagrama 1.

La inteligencia clínica en el sector de la salud se orienta en dos fuentes de información de entrada (externos e internos) y procesamiento de la información en sistemas computarizados cuyas técnicas de desarrollo siguen un orden metodológico personalizado y estructurado a partir de enfoque de algoritmos múltiples o no lineales.

El tratamiento matemático permite estructurar la información de salida cuyo objetivo se adecua a las necesidades de análisis propias de cada tipo de paciente, análisis de diagnóstico de escenarios múltiples, administración de riesgo asociado a cada tipo de tratamiento y planteamiento de elección de la atención apropiada para cada cliente (Murdoch & Detsky, 2013).

Las soluciones cuantitativas al alcance de la información y estructuración de los datos han generado el interés de las ciencias exactas por el desarrollo de múltiples algoritmos enfocados en la identificación y toma de decisiones informativas en el sector de la salud.

Así lo establece Übeyli (2007), en su investigación y generación de propuestas de algoritmos en las decisiones clínicas de dos áreas de estudio específicas para una clasificación de paciente específica: Personas con diabetes y cáncer de mama. El autor plantea una propuesta de solución enfocada en Redes neuronales artificiales (*Artificial neural networks*)

DIAGRAMA 1
Ejemplo de recursos principales para la Inteligencia Clínica



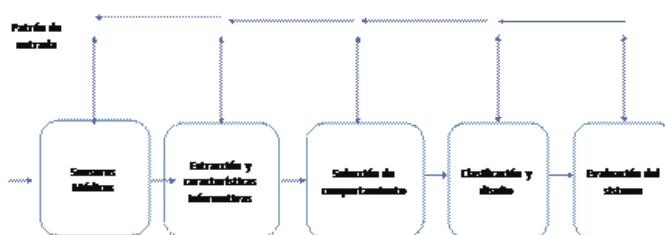
Giacalone & Scippacercola (2016)

-ANNs), software computacional inspirado en neuronas biológicas y análisis de su comportamiento, cuya característica principal se centra en la capacidad de solución óptima más próxima, en donde las redes neuronales pueden combinar información de diferente naturaleza en un único sistema, por ejemplo, protocolos clínicos, datos de laboratorio, señales e imágenes, y lograr como resultado un sistema integrado de diagnóstico, (Diagrama 1).

Una de las aplicaciones principales de Inteligencia clínica y expuestas por (Übeyli, 2007) plantea el desarrollo integrado de sistemas de información de cinco grandes grupos de modelos cuantitativos: redes neuronales combinadas (*Combined Neural Networks - CNNs*), mezcla de expertos (*Mixture of Experts - ME*), redes neuronales probabilísticas (*Probabilistic Neural Networks- PNN*), redes neuronales recurrentes (*Recurrent Neural Networks -RNN*) y máquinas de vectores de soporte (*Support Vector Machines -SVM*) y redes neuronales de perceptrón multicapa(*Multilayer Perceptron Neural Networks- MLPNNs*), siendo este último el sistema cuantitativo utilizado como herramienta de estudio en la presente investigación.

Como fase de entrada, se establecen los datos fuentes, también conocidos como suministro de información, de característica no independiente y por el contrario con un alto grado de correlación entre ellos, en donde modificaciones en los resultados pueden generar cambios en las configuraciones en el sistema. Este tipo de información es ilustrada en el diagrama 2.

DIAGRAMA 2
Etapas básicas involucradas en el diseño de un sistema de clasificación



Giacalone & Scippacercola (2016)

Los datos de entrada en el análisis de la información se presentan a seguir. Un primer comportamiento que analiza el autor son las informaciones en relación a la descripción de los atributos de análisis (características informativas) del caso de estudio de los pacientes con diagnóstico de diabetes (tabla 3.). En este tipo de análisis, se exponen las principales características de los pacientes que sufren de diabetes (Tipo I: diabetes insulino-dependiente y tipo II: diabetes no insulino-dependiente).

Algunas características de la población de estudio identifican resaltan que la población femenina se inclina por la característica de diabetes tipo II. Se evalúan dos condiciones: la primera es si la comida precipita la condición diabética y el segundo es el tipo de si la condición del tipo de comida impulsa la insulino dependencia o no.

En el caso del presente estudio, se estudia la base de datos Pima database (*Neunet Pro - Neural Network Software*, 2019), como referente en la asociación y estudio del tipo de familias de sujetos (en este caso madre gestantes) y reconocimiento del agente patógeno por primera vez durante el embarazo y mayor riesgo de desarrollo de estos agentes.

Un segundo elemento analizado es la aparición de cáncer de mama. Investigaciones recientes han demostrado que la aparición de este tipo de células cancerígenas se debe a diversos factores de riesgo (edad, factores genéticos, antecedentes familiares, obesidad, entre otros), sin embargo, se desconoce que causa más cáncer de seno o cómo este tipo de factores hacen que estas células se vuelvan cancerosas.

Para analizar el comportamiento anterior, se analizan muestras de la base de datos de uno de los hospitales de Wisconsin (Wolberg & Mangasarian,

TABLA 1
Base de datos Pima, Instituto de Diabetes

Tabla 1
Base de datos Pima, Instituto de Diabetes

Número de atributo	Descripción	Promedio	Desviación Estándar
1	Número de veces embarazada	3,8	3,4
2	Concentración de glucosa en plasma a 2h - Prueba de tolerancia a la glucosa	120,9	32
3	Presión arterial diastólica (mmHg)	69,1	19,4
4	Grosor del pliegue de tríceps (mm)	20,5	16
5	Insulina sérica de 2 horas (μ U=ml)	79,8	115,2
6	Índice de masa corporal	32	7,9
7	Probabilidad de diabetes (ancestral)	0,5	0,3
8	Edad (años)	33,2	11,8

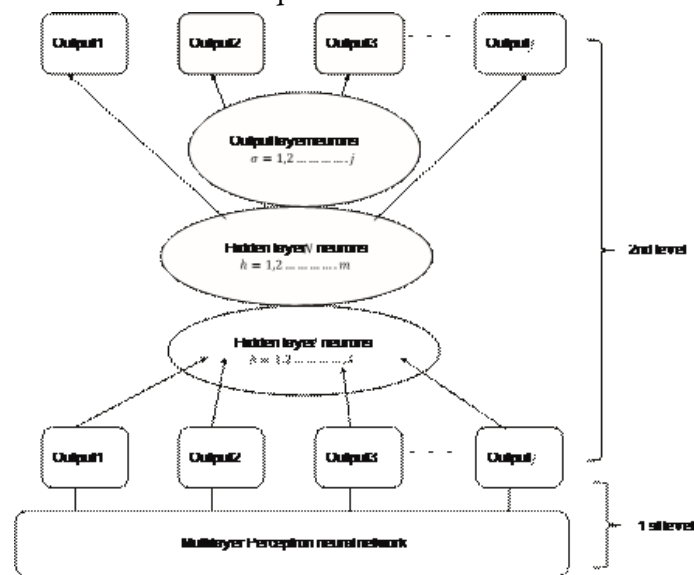
N = 768 observaciones, 268 diabéticos y 500 no diabéticos.

Fuente: Giacalone & Scippacercola (2016)

2009), en relación al tejido mamario y las características de información nuclear, con nueve atributos diferentes. Lo anterior es expuesto en la tabla 1 y 2.

Para el análisis de las informaciones Übeyli (2007), identifica cinco categorías de sistemas de diagnóstico automatizados, *Support vector machine (SVM)*, *Probabilistic Neural Network (PNN)*, *Multilayer Perceptron Neural Network (MLPNN)*, *Expectation Maximization (ME)* y *CNN Models (ConvNet)*.

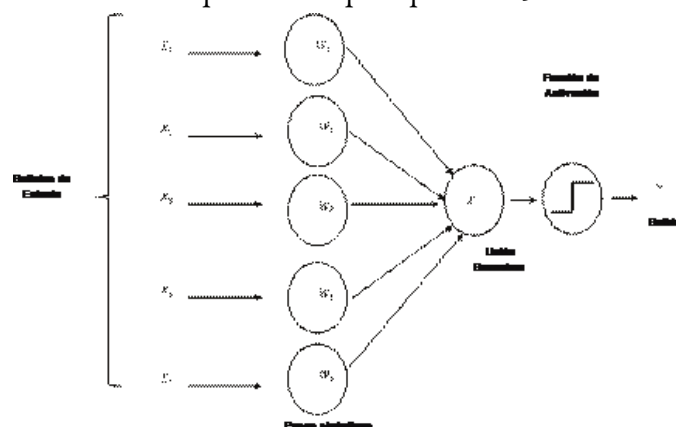
DIAGRAMA 3
Arquitectura CNN



Se realiza una comparación de los resultados obtenidos en cada uno de los sistemas enfocados en el diagnóstico exacto de la población de estudio analizada. Un ejemplo de la arquitectura del sistema de diagnóstico de la presente investigación se presenta en el diagrama 3.

Este tipo de modelo de diagnóstico automatizado está enfocado en el denominado agente de “generalización apilada” (Multilayer perceptron neural network), en donde la arquitectura sigue un orden estructurado de entrada de características informativas (1st level) y obtención de la información deseada de salida Output (diagrama 3).

DIAGRAMA 4
Proceso Sináptico de un perceptrón de 5 entradas



Giacalone & Scippacercola (2016)

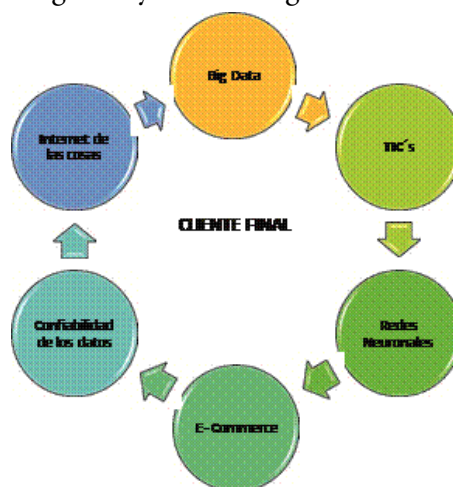
Lo anterior es obtenido a partir de las informaciones de redes neuronales de perceptrones, los cuales se caracterizan por ser redes neuronales artificiales de inferencia cuyo objetivo es la transformación de agentes de entrada de información (señales de entrada denominadas “dendritas” o “sensores de obtención de la información”) (, , ,,) con una asignación de peso (, , ,,) proceso de transformación de la información y desarrollo de un algoritmo binario de decisión (denominado proceso de “sinapsis” o Unión sumatoria) y canales de información de salida conocidos como axones () (diagrama 4).

El proceso de conversión de los axones de salida () genera, como resultado las denominadas redes neuronales ocultas (hidden layer neurons) , esta fase de redes neuronales ocultas es lo descrito por BuendíaRodríguez et al, (2002), como el proceso de simulación y adaptación de las necesidades de simulación deseados a nivel informático, comúnmente utilizado en el campo de la Ingeniería partir de los procesos de simulación y análisis del comportamiento neuronal en pacientes (2nd level), (Pennington et al, 2018)

3. Manufactura de servicios enfocados en la industria 4.0: Big data integrado al sector logístico

En el mundo cada vez más globalizado y dinámico, los entornos de los mercados se mueven de manera vertiginosa y cada vez más acelerados, exigiendo que el tiempo y la información de sus clientes, su entorno y su mercancía tengan que estar disponible en tiempo real para cada uno de los *stakeholder* que participan en esta cadena (Bastos & Gallego, 2016).

DIAGRAMA 5
Big Data y las tecnologías asociadas



Giacalone & Scippacercola (2016)

Una de las herramientas más poderosas utilizada en la logística hoy en día, es la Big data, pues este le permite captar grandes volúmenes de información a las empresas, procesarla, clasificarla, direccionarla para entender la dinámica de los clientes internos y externos de una compañía, sus preferencias, sus hábitos, sus tiempos de compra y otra información que determina como deben actuar las compañías. Ahora bien, la importancia del *Big data* también se da debido a que esta trabaja de la mano con las siguientes tecnologías como lo muestra la gráfica (diagrama 5):

4. INDUSTRIA 4.0 EN LA MANUFACTURA DE SERVICIOS FINANCIERA- TENDENCIAS Y REPERCUSIONES EN APLICACIONES BIG DATA.

Como aspecto a este último sector a evaluar, se hace énfasis inicial en el marco contextual que rodea a gran parte de los procesos productivos y de servicios que se desarrollan a nivel global, y el cual está relacionado con la economía del conocimiento.

De acuerdo con lo anterior es importante resaltar como la aplicación de las 5 Vs (Volumen, Variedad, Velocidad, Veracidad y Valor) (diagrama 6) es de gran utilidad y pertinencia para el sector, dado el gran Volumen de información que se maneja, la alta Variabilidad de esta información y la Velocidad con la que se requiere sea procesada, con el fin de garantizar altos estándares de calidad con respecto los que se refiere a los tiempos de atención y procesamiento de los requerimientos de los clientes, sin perder nunca de vista la Veracidad y confiabilidad de esta información con fines de minimizar los riesgos transaccionales asociados, generando de esta forma Valor a partir de la información.

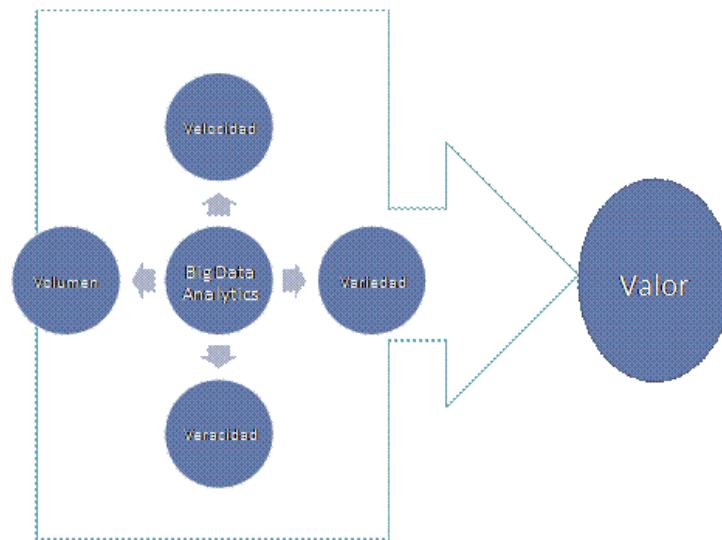


DIAGRAMA 6
Big data analytics y las 5 Vs
Tabares & Hernández (2014)

de las fortalezas competitivas que se pueden desarrollar a partir de la aplicación del modelo Big Data dentro del sector financiero y de acuerdo con (Tabares & Hernández, 2014) se presentan y resumen finalmente en la figura 10, a partir de la cual se identifica la potencialidad de obtener grandes beneficios por parte del sector, los cuales van hasta la transformación de una economía y la llegada de una nueva ola de crecimiento productivo para una región. Las ventajas que se pueden obtener a través del aprovechamiento de Big Data se ilustran en la Gráfico 1, e incluyen; el incremento de la eficiencia operativa, mantener informada a la dirección estratégica, mejora del servicio al cliente, entre otras.

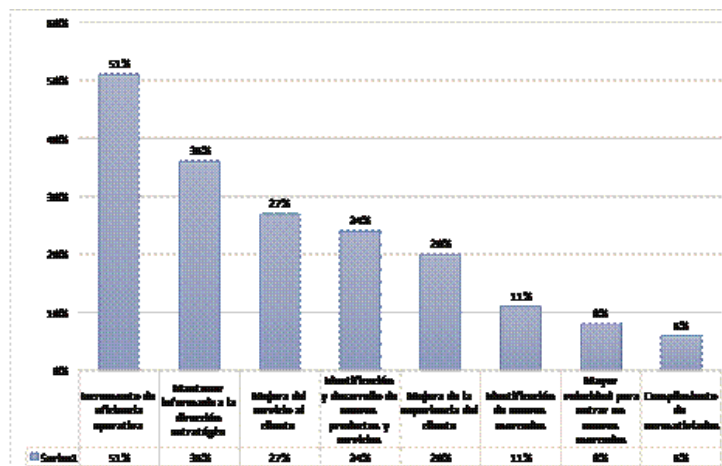


GRÁFICO 1
Ventajas que proporciona el análisis Big Data
Tabares & Hernández (2014)

5. CONCLUSIONES

Una de las implementaciones destacadas de la industria 4.0 se presenta en las aplicaciones del Big Data en el sector de la salud y su relación a la demanda de servicios del cliente. La interpretación e integración de la información de los diferentes servicios (exámenes médicos, citas y consultas, etc.) y el proceso de sinapsis

en la transformación e interpretación de la información a partir de sistemas de diagnóstico automatizados permite a los profesionales del sector generar respuestas con altos niveles de eficiencia y exactitud en la toma de decisiones médicas para el tratamiento de los pacientes.

La integración de hardware y software en el campo del Big Data se hace indispensable en pro de la mejora de la calidad de servicio de teleasistencia en los pacientes. Cuando se vincula redes de información estructuradas (*cluster hadoop*) en el tratamiento y asistencia sanitaria de pacientes, se hace fundamental el conocimiento de la información analizada (*Knowledge Discovery in Databases – KDD*) que permitan la obtención de información de entrada para su correcto análisis

REFERENCIAS BIBLIOGRÁFICAS

- Abugabah, A., Nizamuddin, N., & Abuqabbah, A. (2020). A review of challenges and barriers implementing RFID technology in the Healthcare sector. *Procedia Computer Science*, 170, 1003–1010. <https://doi.org/10.1016/j.procs.2020.03.094>
- Bastos, K. L., & Gallego, A. (2016). ¿Qué está cambiando el rumbo de la logística? Cinco tendencias logísticas en la cadena de suministro. *Mercatec*, 51, 19–27.
- Buendía-Rodríguez, E., Vargas-Pérez, E., Leyva-Ovalle, Á., y TerrazasDomínguez, S. (2002). Aplicación de redes neuronales artificiales y técnicas sig para la predicción de coberturas forestales. *Revista Chapingo. Serie Ciencias Forestales y del Ambiente*, 8(1), 31–37.
- Din, I. U., Guizani, M., Rodrigues, J. J. P. C., Hassan, S., & Korotaev, V. V. (2019). Machine learning in the Internet of Things: Designed techniques for smart cities. *Future Generation Computer Systems*, 100, 826–843. <https://doi.org/10.1016/j.future.2019.04.017>
- Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210, 15–26. <https://doi.org/10.1016/j.ijpe.2019.01.004>
- Giacalone, M., & Scippacercola, S. (2016). *Quantitative Methods Inquires I Big Data*. Issues And An Overview In Some Strategic Sectors.
- Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model. *Procedia CIRP*, 79, 403–408. <https://doi.org/10.1016/j.procir.2019.02.106>
- Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *In JAMA - Journal of the American Medical Association*, 309(13), 1351–1352. <https://doi.org/10.1001/jama.2013.393>
- Patil, H. K., & Seshadri, R. (2014). Big data security and privacy issues in healthcare. *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014*, 762–765. <https://doi.org/10.1109/BigDataCongress.2014.112>
- Pennington, J., Brain, G., Worah, P., & Research, G. (2018). *The Spectrum of the Fisher Information Matrix of a Single-Hidden-Layer Neural Network*.
- Perera, C., Ranjan, R., Wang, L., y Ullah, S. (2014). Big Data Privacy in the Internet of Things Era. *IT Professional*, 17(3). <https://doi.org/10.1109/MITP.2015.34>
- Shakhovska, N., Boyko, N., Zasoba, Y., & Benova, E. (2019). Big data processing technologies in distributed information systems. *Procedia Computer Science*, 160, 561–566. <https://doi.org/10.1016/j.procs.2019.11.047>
- Tabares, L. F., & Hernández, J. F. (2014). *Big Data Analytics: Oportunidades, Retos y Tendencias*.
- Übeyli, E. D. (2007). Comparison of different classification algorithms in clinical decision-making. *Expert Systems*, 24(1), 17–31. <https://doi.org/10.1111/j.1468-0394.2007.00418.x>
- Vieira, A. A., Dias, L. M., Santos, M. Y., Pereira, G. A., & Oliveira, J. A. (2020). On the use of simulation as a Big Data semantic validator for supply chain management. *Simulation Modelling Practice and Theory*, 98, 101985. <https://doi.org/10.1016/j.simpat.2019.101985>

- Wolberg, W. H., & Mangasarian, O. L. (2009). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the National Academy of Sciences of the United States of America*, 87(23), 9193–9196. <https://doi.org/10.1073/pnas.87.23.9193>
- Zhang, D. (2018). *Big Data Security and Privacy Protection*. <https://doi.org/10.2991/icmcs-18.2018.56>