

Herramientas de análisis de grandes volúmenes de datos para la toma de decisiones empresariales y fidelización de los clientes

UMAQUINGA, Ana*†, SALTOS, Tatiana y PELUFFO, Diego

Recibido Abril 22, 2017; Aceptado Junio 9, 2017

Resumen

El crecimiento acelerado de la competencia entre las empresas a nivel global es innegable. Los nuevos retos para generar valor, descubrir tendencias y preferencias de consumo de clientes exigen a las empresas contar con herramientas tecnológicas de mayor alcance y eficiencia que permitan ser apoyo en la toma de decisiones en tiempo real o en el menor tiempo posible con el fin de crear ventaja competitiva y mantenerse en el mercado. Herramientas tecnológicas como Big data, Data mining entre otras, pretenden obtener nuevo conocimiento, patrones, predicciones para una mejor comprensión y precisión de las necesidades del cliente, que no sería posible con el uso de herramientas convencionales. De ahí el trabajo en equipo del área tecnológica y del negocio requiere de este conocimiento de valor para la definición de estrategias encaminadas a capitalizar dicho conocimiento originado de manera interna y externa especialmente de las redes sociales. El presente artículo es de tipo descriptivo, exploratorio, y documental, se realiza un análisis de Big data y minería de datos a nivel empresarial, sus tópicos de uso en la toma de decisiones y búsqueda de la fidelización de sus clientes, así como sus aspectos de valor y retos. En consecuencia, los resultados evidencian la relación favorable y eficiente de la aplicación de estas tecnologías en la toma de decisiones empresariales, por lo que, surge la necesidad de un mayor número de investigaciones de este tipo.

Datos masivos, Data mining, Big data, herramientas para toma de decisiones, ventaja competitiva

Abstract

The accelerated growth of competition among companies globally is undeniable. The new challenges to generate value, discover trends and preferences of customers demand companies to have technological tools of greater reach and efficiency that allow to be support in the decision making in real time or in the shortest possible time in order to create competitive advantage and stay in the market. Technological tools like Data mining, Big data, among others, aim to obtain new knowledge, patterns, predictions for a better understanding and precision of customer needs, which would not be possible with the use of conventional tools. Hence teamwork in the area of technology and business requires this knowledge of value for the definition of strategies aimed at capitalizing on this knowledge originated in an internal and external way especially of social networks. The present article is descriptive, exploratory, and documentary, perform an analysis of the technological tools used at the business level and topics of use in the decision making and search of the loyalty of its clients, as well as its value aspects and challenges. Consequently, the results show the favorable and efficient relationship of the application of these technologies in business decision-making, therefore, the need arises for a greater number of future research of this type.

Information systems, server, Internet, web, mobile devices

Citación: UMAQUINGA, Ana, SALTOS, Tatiana y PELUFFO, Diego. Herramientas de análisis de grandes volúmenes de datos para la toma de decisiones empresariales y fidelización de los clientes. Revista de Estrategias del Desarrollo Empresarial. 2017, 3-8: 31-39

*Correspondencia al Autor (Correo Electrónico: acumaquina@utn.edu.ec)

† Investigador contribuyendo como primer autor.

Introducción

El crecimiento exponencial de los datos estructurados, semiestructurados y no estructurados como videos, imágenes, correos electrónicos, entre otros, gracias a las nuevas tecnologías y a la frecuencia de su uso (Camargo-Vega, Camargo-Ortega, & Joyanes-Aguilar, 2014), han impactado trascendentalmente a la sociedad, ciencia, tecnología y economía (IBM & Martins, 2017a). El mantener e incrementar la permanencia en el mercado de las empresas es cada vez un reto mayor, por una parte, la transformación digital está redefiniendo el futuro hacia la era de los llamados negocios cognitivos y hacia una mayor incorporación de nuevas formas de hacer negocios (IBM, 2017b), mientras que por otra, estos nuevos retos y desafíos requieren de la misma tecnología o de una mayor con la combinación de varias áreas de la ciencia para proporcionar el valor agregado a los clientes, que no puede ser obtenida utilizando los informes y herramientas tradicionales (IBM, 2012).

La generación del Milenio, también llamada la Generación Z requiere de una mayor, intuitiva, relevante y atractiva tecnología de consumo (IBM, 2017a). Es así, que áreas como el Big data, minería de datos (*Data mining*), inteligencia artificial, aprendizaje de máquina (*Machine learning*), cómputo cognitivo buscan ser la respuesta a las demandas actuales y futuras del mercado.

El presente artículo procura ser de aporte a la comunidad de investigadores tanto del área tecnológica, económica y afines, puesto que, presenta una sinergia de enfoques y conocimientos en la búsqueda de soluciones para lograr la fidelización de los clientes.

El resto del documento se organiza como se explica a continuación En la sección 2 se presenta en la metodología las herramientas tecnológicas como Big data y los enfoques de uso de las empresas para esta tecnología, así como de las herramientas mayormente utilizadas, mientras que, para minería de datos se revisa en el aspecto técnico sobre como trabaja para descubrir nuevo conocimiento. En la sección 3 se presenta los principales resultados obtenidos de esta revisión de literatura, mientras que en la sección 4 las conclusiones del presente estudio.

Metodología

Con la finalidad de analizar los aspectos planteados en este estudio, la metodología de investigación es de tipo exploratorio, descriptivo y documental. Uno de los modelos que utiliza el área de Gestión de la información y conocimiento es la cadena de Datos-Información-Conocimiento (*DIC*), también llamada jerarquía de la Información o pirámide del conocimiento (Guillén, Ayuso, Paniagua, & Cadenas, 2015). A continuación, se establece las siguientes definiciones fundamentales que se usarán a lo largo de este trabajo: **Dato:** Son señales o patrones sin significado ni contexto (Schreiber & Akkermans, 2000) (Guillén et al., 2015). **Información:** Se define como datos con significado en un contexto, es decir, la salida del proceso de interpretación y la entrada al proceso de aprendizaje (Aamodt & Nygard, 1995) (Guillén et al., 2015) (Schreiber & Akkermans, 2000). **Conocimiento:** “Es la información aprendida que se incorpora como recurso activo de razonamiento interiorizado dentro del proceso de decisión de un agente natural o artificial” (Guillén et al., 2015).

La marcada diferencia entre dato, información y conocimiento va desde su menor a mayor valor e importancia respectivamente, y de un mayor a menor tamaño.

Los “datos masivos” al introducirse en todos los campos de la actividad humana, demuestran una considerable transformación en el tratamiento de información e innovación ya que evidencian su valor científico, societario hasta convertirse en una fuente de valor económico (Mayer-Schönberger & Cukier, 2013) Áreas como Big data, aprendizaje automático, minería de datos, entre otros, requieren de métodos capaces de extraer conocimiento útil y valioso de los grandes volúmenes de datos, utilizando el descubrimiento de conocimiento en bases de datos (KDD) para una mejor toma de decisiones a nivel gerencial con el fin de tener ventaja competitiva favoreciendo de esta manera, la operatividad del negocio, la reducción de tiempos de respuesta en el análisis de datos para obtener la personalización de productos y servicios a clientes actuales, posibles y potenciales.

Big data como soporte para la toma de decisiones

Big data se enfoca en capturar, curar, almacenar, buscar, compartir, transferir, analizar, visualizar millones de datos (Hammer, He, & Martinetz, 2014). Se caracteriza por 4 dimensiones principales: volumen (petabytes de información (Márquez & Lev, 2016), presenta un incremento considerable de volumen por el uso del Internet de las cosas IoT (Gartner, 2015)), variedad (datos de diferente tipo, entre ellos: imágenes, lenguaje natural y sonidos encontrados en libros, correos electrónicos, tweets, blogs, imágenes y archivos de audio y video (IBM & Martins, 2017b)), velocidad, valor (beneficios, reducción de costos) (Curto, 2012), adicionalmente se integra 2 nuevas dimensiones: veracidad (fiabilidad de la información (Umaquina, Peluffo, Alvarado P., & Cabrera A., 2016)) y visualización (presentación legible de la información (Sivarajah, Kamal, Irani, & Weerakkody, 2016)) (Camargo-Vega et al., 2014).

Entre los tipos de análisis que utiliza se encuentra: Descriptivo (analiza datos tanto históricos como actuales para identificar las tendencias y relaciones), predictivo (aprendizaje no supervisado, estudia la probabilidad de ocurrencia), prescriptivo (análisis de una decisión con sus variables para identificar la probabilidad que se produzca), causal (busca los causales de problemas para establecer propuestas de solución) (Rosero, Rosero, Peluffo, & Beltrán, 2016).

El área de visualización de información en Big data (denominada Analítica Visual, VA, Data Vis, VI) busca apoyar al usuario en la exploración, entendimiento, y análisis de datos por medio de la exploración visual, progresiva e iterativa (Liu, Cui, Wu, & Liu, 2014). (Umaquina et al., 2016) establece un acercamiento e identificación de técnicas aplicadas en herramientas Open Source y comerciales para visualización de información de Big Data: 58 técnicas de visualización: Geometric Projection, Icon-Based, Pixel-oriented, Hierarchical, Graph-Based, Hybrid, Radial coordinate visualization, visual structures, interaction, tabular data, Graphs, Geospatial, Tree, visualization techniques, entre otras y 31 herramientas de software: 21 herramientas de tipo open Source y 10 de tipo comercial con el fin de analizar la tendencia de uso en los usuarios finales.

No obstante, (Redmond & Wilson, 2012) citado en (Camargo-Vega et al., 2014) señala como debilidades:

“Falta de personal y habilidades del recurso humano (46%), la dificultad en la arquitectura de un sistema de análisis de Big Data (33%), problemas con el Big Data utilizable para los usuarios finales (22%), la falta de patrocinio empresarial (38%) y la deficiencia de un argumento empresarial convincente (28%), la carencia de análisis de bases de datos (32%), problemas de escalabilidad de Big Data (23%), rapidez en las consultas (22%) y dificultad para cargar los datos lo suficientemente rápido (21%), entre otros”

Es preciso mencionar que de 352 encuestados por (tdwi, 2016) consideran extremadamente importante la implementación de Big data para los siguiente requerimientos detallados en la Tabla 1:

Descripción	%
• Información empresarial más precisa	57%
• Comprender a los clientes	49%
• Predecir comportamiento	44%
• Mejorar el rendimiento / prácticas / procesos del negocio	40%
• Diferenciador competitivo	32%
• Impulsar la eficiencia operativa	31%
• Responder más rápida al cambio de negocios	30%
• Identificar los riesgos y / o reducir el fraude	28%
• Monetizar nuestros análisis	22%
• Desarrollar de nuevas aplicaciones basadas en datos	19%

Tabla 1 Áreas importantes para implementar Big data

Fuente (tdwi, 2016)

(Intel, 2012) identifica los requerimientos de análisis de tipos de datos como se detalla en la Tabla 2.

Origen de datos	% de requerimiento de análisis
• Documentos	84%
• Transacciones empresariales base de datos	82%
• E-mail	84%
• Datos de imágenes	68%
• Sensor o dispositivo de datos	57%
• Búsquedas indexadas de internet	57%
• Weblogs	55%
• Social media	54%
• Conversaciones telefónicas	52%
• Videos	52%
• Imágenes	46%
• Clic	42%
• Otros	2%

Tabla 2 Análisis de requerimientos enfocado de tipos de datos

Fuente (Intel, 2012)

Mientras que las funciones principales de impacto comercial por tipo de industria se presentan en la Tabla 3:

Industria	Función principal de impacto comercial
Al por menor	<ul style="list-style-type: none"> • Venta cruzada
	<ul style="list-style-type: none"> • Marketing basado en la ubicación
	<ul style="list-style-type: none"> • Análisis de sentimientos
	<ul style="list-style-type: none"> • Optimización de la cadena de suministro
Manufactura	<ul style="list-style-type: none"> • Información en tiempo real de la cadena de suministro
	<ul style="list-style-type: none"> • Análisis de proceso usando red de sensores
	<ul style="list-style-type: none"> • Optimización de distribución
	<ul style="list-style-type: none"> • Mantenimiento predictivo
Telecomunicaciones	<ul style="list-style-type: none"> • Optimización de red
	<ul style="list-style-type: none"> • Prevención Churn
Sector público	<ul style="list-style-type: none"> • Incremento de la productividad
	<ul style="list-style-type: none"> • Creación de transparencia
	<ul style="list-style-type: none"> • Bioinformática
	<ul style="list-style-type: none"> • Predicción/Simulación
	<ul style="list-style-type: none"> • Medicina personalizada
	<ul style="list-style-type: none"> • Evaluación de eficacia de fármacos

Tabla 3 Función principal de impacto comercial por tipo de industria

Fuente (Márquez & Lev, 2016)

En el informe de mejores prácticas de Transforming Data with Intelligence (denominado *TDWI*) (tdwi, 2016) establece como una posible predicción de uso de herramientas tecnológicas en Big data para el año 2017 representado en la Tabla 3:

Herramienta/Tecnología/Lenguaje	% predicción de uso
<ul style="list-style-type: none"> • Hadoop 	41%
<ul style="list-style-type: none"> • R 	38%
<ul style="list-style-type: none"> • Data Warehouse 	35%
<ul style="list-style-type: none"> • Spark 	22%
<ul style="list-style-type: none"> • Data Virtualization 	21%
<ul style="list-style-type: none"> • In-memory technologies 	20%
<ul style="list-style-type: none"> • The hybrid cloud 	17%
<ul style="list-style-type: none"> • Python 	17%
<ul style="list-style-type: none"> • Real-time database capabilities 	13%
<ul style="list-style-type: none"> • Massively parallel processing platforms 	12%
<ul style="list-style-type: none"> • Event stream processing engines 	11%
<ul style="list-style-type: none"> • The private cloud 	10%
<ul style="list-style-type: none"> • The public cloud 	7%
<ul style="list-style-type: none"> • Content management technologies 	6%
<ul style="list-style-type: none"> • Third-party data services 	6%
<ul style="list-style-type: none"> • Kafka 	4%

Tabla 4 Predicción de uso de herramienta/ tecnología/ lenguaje en Big data

Fuente (tdwi, 2016)

Minería de datos

También llamada explotación de información (Moine, Haedo, & Gordillo, 2011) de grandes volúmenes de datos, procura encontrar información oculta, cuyo origen son bases de datos operacionales o bodegas de datos conocidos como (denominado Datawarehouse), se apoya en la estadística, computación y matemática.

Con el uso del análisis predictivo, y tecnología requerida permite para construir sistemas que permitan aprender de los datos para identificar patrones y predecir resultados futuros con mínima intervención humana, es decir, que el computador aprende enfoques supervisados o no supervisados (tdwi, 2016).

Una de las principales áreas de enfoque de la minería de datos es el análisis de problemas sociales, así como, la identificación de características de la población, estudiantes, segregación residencial y académico, desempleo (Franco-Arcega, Franco-Sánchez, Castro-Espinoza, & García-Islas, 2014)

(Tan, Steinbach, & Kumar, 2005) establece 5 aspectos clave para su proceso: exploración de datos (para establecer sobre el preprocesamiento más apropiado y técnicas de análisis de datos), clasificación (técnica de aprendizaje automático que cataloga los datos en grupos predefinidos), agrupamiento/clustering (técnica de aprendizaje no supervisada, permite analizar patrones significativos sin supervisión humana intervención.), asociación (descubre la probabilidad de que ocurra un evento específico como resultado directo de otro) y análisis de series de tiempo.

A continuación, se presenta las principales técnicas que utiliza la minería de datos, Tabla 4:

Tipo de Modelo	Subtipos
Dirigido por la teoría. Técnicas predictivas	<ul style="list-style-type: none"> Modelos de regresión
	<ul style="list-style-type: none"> Análisis de varianza
	<ul style="list-style-type: none"> Series temporales
	<ul style="list-style-type: none"> Análisis discriminante
	<ul style="list-style-type: none"> Arboles de decisión
	<ul style="list-style-type: none"> Redes neuronales
Dirigido por los datos. Técnicas descriptivas	<ul style="list-style-type: none"> Análisis clúster
	<ul style="list-style-type: none"> Escalamiento multidimensional
	<ul style="list-style-type: none"> Escalamiento óptimo
	<ul style="list-style-type: none"> Reducción de dimensión
	<ul style="list-style-type: none"> Segmentación
Técnicas auxiliares	<ul style="list-style-type: none"> Proceso analítico de transacciones (OLAP)
	<ul style="list-style-type: none"> Reporting

Tabla 5 Principales técnicas de minería de datos

Fuente (Marqués, 2014)

Varias son las propuestas metodológicas como SEMMA correspondiente a las cinco fases básicas del proceso: Sample (Muestreo), Explore (Exploración), Modify (Modificación), Model (Modelado), Assess (Valoración)(Moine et al., 2011) que organiza en herramientas llamadas nodos para el procedo de Minería de datos.

Resultados

Por lo tanto, se puede establecer que la combinación y sinergia de áreas matemáticas, estadística y computación permite la obtención de patrones, segmentación ajustada de proceso, para entender el comportamiento de los clientes (IBM, 2017c) y lograr óptimas experiencias e incremento de la fidelidad de los clientes, así como la creación y seguimiento de campañas personalizadas en offline y online (IBM, 2017c) permite elevar el valor monetario de los datos en la obtención de conocimiento a favor de la empresa y el cliente.

Los usos más requeridos para el análisis en Big data en la industria se encuentran: documentos, transacciones empresariales/ base de datos, e-mails.

Además, como funciones de impacto e innovación comercial, se observa la aparición del estudio de análisis de sentimientos, que posibilitaría el precisar y segmentar los requerimientos de los clientes al disponer de una mayor disponibilidad de datos e información, en tiempo real desde las redes sociales.

La predicción de uso de herramienta/ tecnología/ lenguaje de Big data prioriza a Hadoop, R, Data Warehouse, lo que posibilitaría establecer nuevos requerimientos de profesionales en este tipo de herramientas.

En el área de la minería de datos, existe diversidad de técnicas de uso entre ellos los modelos predictivos, descriptivos, auxiliares que pretenden apoyar en la extracción de conocimiento de valor.

Conclusiones

Esta investigación presenta una revisión descriptiva del área de Big Data y Minería de datos enfocado a empresa, con un recorrido por principales características, importancia, retos, oportunidades, así como su pronóstico de uso, las plataformas y su porcentaje de uso actual, con el fin de presentar a la comunidad empresarial y de investigadores una guía inicial y descriptiva sobre la importancia del uso de estas nuevas tecnologías y propuestas en beneficio de la empresa y el abanico de clientes.

Agradecimiento

Las autoras agradecen a la Universidad Técnica del Norte, a la Carrera de Ingeniería en electrónica y Redes de Comunicación de la ciudad de Ibarra – Ecuador.

Referencias

- Aamodt, A., & Nygard, M. (1995). Different Roles and Mutual Dependencies of Data, Information, and Knowledge—an AI Perspective on Their Integration. *Data Knowl. Eng.*, 16(3), 191–222. [https://doi.org/10.1016/0169-023X\(95\)00017-M](https://doi.org/10.1016/0169-023X(95)00017-M)
- Camargo-Vega, J. J., Camargo-Ortega, J. F., & Joyanes-Aguilar, L. (2014). Knowing the Big Data. *Revista Facultad de Ingeniería*, 24(38), 63-77.
- Curto, J. (2012). *Resumen Ejecutivo - Big Data :un mercado emergente*. Recuperado a partir de <http://www.diarioabierto.es/wp-content/uploads/2012/06/Resumen-Ejecutivo-IDC-Big-Data.pdf>

Franco-Arcega, A., Franco-Sánchez, K. D., Castro-Espinoza, F. A., & García-Islas, L. H. (2014). Data Mining for Discovering Patterns in Migration. En *Nature-Inspired Computation and Machine Learning* (pp. 285-295). Springer, Cham. https://doi.org/10.1007/978-3-319-13650-9_26

Gartner. (2015). Gartner Says 6.4 Billion Connected. Recuperado 20 de junio de 2017, a partir de <http://www.gartner.com/newsroom/id/3165317>

Guillén, M. A., Ayuso, B. L., Paniagua, E., & Cadenas, J. M. (2015). Una revisión de la Cadena Datos-Información-Conocimiento desde el Pragmatismo de Peirce. *Documentación de las Ciencias de la Información*, 38(0), 153-177. https://doi.org/10.5209/rev_DCIN.2015.v38.50814

Hammer, B., He, H., & Martinetz, T. (2014). Learning and modeling big data. En *ESANN* (pp. 343-352).

IBM. (2012, noviembre 19). Dónde comenzar con la minería de datos en distribución mayorista. Recuperado 23 de octubre de 2017, a partir de <http://www.ibm.com/developerworks/ssa/industry/library/ba-data-mining-wholesale/index.html>

IBM. (2017a, enero 30). Generación Z: los consumidores del futuro. Recuperado 24 de octubre de 2017, a partir de <https://www.ibm.com/blogs/transformacion/2017/01/30/generacion-z-los-consumidores-del-futuro/>

IBM. (2017b, junio 28). La infraestructura en la Era Cognitiva. Recuperado 24 de octubre de 2017, a partir de <https://www.ibm.com/blogs/transformacion/2017/06/28/la-infraestructura-en-la-era-cognitiva/>

IBM. (2017c, octubre 5). Revolución del marketing en las empresas y las más de 5 mil marcas que están triunfando. Recuperado 23 de octubre de 2017, a partir de <https://www.ibm.com/blogs/transformacion/2017/10/05/revolucion-del-marketing-en-las-empresas-las-5-mil-marcas-estan-triunfando/>

IBM, & Martins, A. (2017a, mayo 25). El poder de la Reinención Digital de las empresas en la era cognitiva. Recuperado 24 de octubre de 2017, a partir de <https://www.ibm.com/blogs/transformacion/2017/05/25/poder-la-reinencion-digital-las-empresas-en-la-era-cognitiva/>

IBM, & Martins, A. (2017b, agosto 23). Cómo se construye una empresa cognitiva. Recuperado 24 de octubre de 2017, a partir de <https://www.ibm.com/blogs/transformacion/2017/08/23/se-construye-una-empresa-cognitiva/>

Intel. (2012). Peer Research Big Data Analytics. Recuperado a partir de <https://www.intel.com/content/dam/www/public/us/en/documents/reports/data-insights-peer-research-report.pdf>

Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. *The Visual Computer*, 30(12), 1373-1393. <https://doi.org/10.1007/s00371-013-0892-3>

Marqués, M. P. (2014). *Minería de datos a través de ejemplos*. RC Libros. Recuperado a partir de <https://books.google.com.ec/books?id=311foAEACAAJ>

Márquez, F. P. G., & Lev, B. (2016). *Big Data Management*. Springer.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data : la revolución de los datos masivos*. Turner.

Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. Presentado en XIII Workshop de Investigadores en Ciencias de la Computación. Recuperado a partir de <http://hdl.handle.net/10915/20034>

Rosero, P., Rosero, E., Peluffo, D., & Beltrán, L. (2016). ARQUITECTURA, SERVICIOS Y APLICACIONES DE Business Intelligence: Revisión de literatura (PDF Download Available). Quito, Ecuador. Recuperado a partir de https://www.researchgate.net/publication/313280378_ARQUITECTURA_SERVICIOS_Y_APLICACIONES_DE_Business_Intelligence_Revision_de_literatura

Schreiber, G. T., & Akkermans, H. (2000). *Knowledge Engineering and Management: The CommonKADS Methodology*. Cambridge, MA, USA: MIT Press.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2016). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*. <https://doi.org/10.1016/j.jbusres.2016.08.001>

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining* (1 edition). Boston: Pearson.

tdwi. (2016). Best Practices Report | Data Science and Big Data: Enterprise Paths to Success. Recuperado a partir de <https://tdwi.org/research/2016/12/best-practices-report-data-science-and-big-data.aspx>

Umaquina, A. C., Peluffo, D., Alvarado P., J. C., & Cabrera A., M. V. (2016). Estudio descriptivo de técnicas aplicadas en herramientas Open Source y comerciales para visualización de información de Big Data. En *Generando Ciencia: Memorias de las I Jornadas Internacionales de Investigación Científica UTN*. Recuperado a partir de https://www.researchgate.net/publication/312217978_Estudio_descriptivo_de_tecnicas_aplicadas_en_herramientas_Open_Source_y_comerciales_para_visualizacion_de_informacion_de_Big_Data