

ARQUITECTURA, SERVICIOS Y APLICACIONES DE *Business Intelligence*: Revisión de literatura

Paul Rosero, Edwin Rosero, Luis Beltran, Diego Peluffo and Carlos Pupiales

Universidad Técnica del Norte

Abstract— *Business Intelligence (BI) es el conjunto de estrategias y herramientas para analizar gran cantidad de volúmenes de datos con el fin de encontrar patrones o tendencias de consumo de las personas y establecer estrategias de negocio, para lograr este objetivo es necesario contar con servicios y aplicaciones como RealTime BI, Social BI, Cloud BI, BI 3.0, Business Analytics y Mobile BI. Todo el proceso de BI es soportado por diferentes análisis que implementan algoritmos de machine learning en grandes volúmenes y diferentes fuentes de datos, considerado como Big Data. En este trabajo se hace relación a las funcionalidades y los requerimientos necesarios de BI desde un concepto inicial hasta conceptos específicos y herramientas para su implementación.*

Keywords— *Big data, Business Intelligence, Business, competitive intelligence*

I. INTRODUCTION

Las empresas continuamente buscan nuevos segmentos de mercado que conlleven a nuevos modelos de negocio, los usos de métodos de análisis de datos ayudan a medir el rendimiento para la toma de decisiones y poder aumentar la eficacia de una estrategia de venta, este proceso define el éxito de un producto o servicio. *Business Intelligence (BI)* permite la gestión de información en el momento adecuado y extraer conocimiento para la toma de decisiones empresariales, siendo la clave del éxito en las organizaciones, se debe invertir en la infraestructura sea física o en la nube para su implementación [1]. BI necesita de algoritmos de aprendizaje de máquina (*machine learning*) que permiten la toma de decisiones y/o de estadística aplicada, como herramienta de pronóstico, de forma que el BI cuente con diferentes puntos de vista y de valor para las organizaciones [2]. De acuerdo con [3], para la implementación de BI es necesario reconocer el tiempo de ejecución de la estrategia, las metas de la organización y los objetivos planteados con un segmento de mercado claro, tomando en cuenta la relación deseada a entablar con el cliente.

Según [4], los nuevos líderes en plataformas BI son Tableau, Qlik y Microsoft Power BI. Las tecnologías de información (TI) se han convertido en un factor indispensable para las organizaciones, junto con la unión de BI permite una eficiente administración de información en los diferentes niveles de la empresa, permitiendo innovar en la creación de nuevas oportunidades de negocio [2]. La infraestructura de un BI debe contar una plataforma BI de administración, una plataforma como servicio (Cloud BI), seguridad y administración de usuarios y conectividad a las fuentes de datos con capacidad de

administración de metadatos, procesos de ETL (extracción, transformación y carga), como una fase de preparación de datos [4].

En este artículo se presenta una revisión sobre las herramientas y los usos de Business Intelligence, con énfasis en las herramientas de aprendizaje automático disponible para sistemas de soporte de decisión, se hace relación a las funcionalidades y los requerimientos necesarios de BI desde un concepto inicial hasta conceptos específicos para su implementación.

El presente trabajo está estructurado de la siguiente manera: En la sección II se habla sobre la evolución de BI, la sección III sobre big data, el capítulo IV sobre los servicios y aplicaciones de BI, en la sección V se documenta el uso de machine learning para BI y finalmente en la sección VI se presenta una breve discusión y comentarios finales.

II. EVOLUCIÓN DE BUSINESS INTELLIGENCE

A. BI 1.0

Es considerado como el principio de BI y es un enfoque centrado en datos usando el modelo cliente servidor con datos mayormente estructurados en sistemas de gestión relacionales y primeros principios de ETL, y sus informes eran basados en intuiciones básicas. Fue concebido para realizar el estudio predictivo entre servidores operativos, y representó el inicio de la visión a las operaciones de negocios específicos, aunque con información restringida (por ejemplo, usando data management y warehousing). Las plataformas más conocidas fueron Microsoft, IBM, Oracle and SAP [5] [6].

B. BI 2.0

Con el acceso a Internet mediante protocolos web como HTTP (protocolo de transferencia de hipertexto) permitió que los procesos de BI sean factibles para diferentes tamaños de empresas, dando lugar al BI 2.0. Simplificando la adquisición de información, fue posible extraer de forma independiente los datos generados sin requerir que el usuario sea experto en TI, proporcionando mayor flexibilidad y escalabilidad en los sistemas, ya que pueden segmentar al cliente por direcciones de red y se recogían datos a través de cookies y registros de servidores. Posteriormente, en el año 2000, BI 2.0 ya se basaba

en los análisis web no estructurados dando cierta complejidad a los administradores de TI [5] [6].

C. BI 3.0

El uso frecuente de BI 2.0 atrajo a procesos de investigación y de industria por el impacto de los resultados, dio lugar a la nueva versión BI 3.0. El número de teléfonos móviles y tabletas superaron al número de computadoras (380 millones), siendo aproximadamente 480 millones, y en el 2020 se prevé que existan 10 mil millones de dispositivos móviles. Esto supone una transformación de la dinámica sociedad y de la forma cómo se almacena, procesa, analiza y visualiza los datos.

BI 3.0 permite explorar en forma de representación de datos interactiva e intuitiva que debe cerrar la brecha entre la empresa y el consumidor para convertirse en un componente central en el esfuerzo continuo para mejorar el negocio mediante la interacción con redes sociales, siendo fácilmente disponibles y multiplataforma [5] [6]. En la figura 1 se observa la evolución de BI junto con las características de cada uno y los diferentes usos.

III. BIG DATA

El crecimiento de datos como música, videos, correos electrónicos o texto no estructurados aumenta la complejidad en las bases de datos. En [7] se define big data en el orden de peta (10^{15}) o zetta (10^{21}) bytes de información, la cual implica un procesamiento complejo que debe ser ejecutado a una gran velocidad por lo cambiante de la información en relación a los tera bytes por día. Las infraseestructuras de TI necesitan un análisis para determinar la veracidad de los datos con el fin de agregar valor la información y lograr un ventaja competitiva. Los gerentes pueden medir y conocer el desarrollo de su empresa, un ejemplo de esto es la reconocida empresa Amazon [8]. Los datos pueden ser estructurados, semi-estructurados, casi estructurados y no estructurados.

Con la manipulación de los datos se pueden hacer pronósticos del comportamiento del cliente. Algunos autores afirman que la calidad de los datos es más importante que su volumen, sin embargo, cuánto mayor es la cantidad de datos estén disponibles, es mejor la definición de un modelo de pronóstico [9]. Normalmente, para formular un modelo es necesario contar con diferentes variables para aplicar análisis estadísticos con un margen de error mínimo. Con la premisa de que un modelo sólo puede generar supuestos y éstos pueden cambiar en el tiempo, los beneficios deben estar enfocados en la reducción del mínimo riesgo, la identificación de fraude y la búsqueda de nuevas oportunidades de negocio [9]. En general, los tipos de análisis se pueden clasificar en: Análisis descriptivo (Sección II-A), análisis predictivo (Sección II-B), análisis prescriptivo (Sección II-C) y análisis causal (Sección II-D).

A. ANÁLISIS DESCRIPTIVO

Se basan en análisis de datos históricos y actuales para determinar las relaciones y tendencias para identificar la acción que debe tomarse [9]. En análisis descriptivo está relacionado en cómo y porqué las personas actúan y de la forma como lo hacen [10], es un aprendizaje supervisado que solo puede clasificar lo sucedido con anterioridad.

B. ANÁLISIS PREDICTIVO

Se analizan los datos del pasado para determinar la probabilidad de ocurrencia de un evento, sirve de guía para la toma de decisiones, pero siempre se basa en una probabilidad [9], es un sistema de aprendizaje no supervisado, que le indica al análisis prescriptivo dónde se encuentra y hacia donde debe dirigirse para evaluar los riesgos desde un punto de vista y tener una visión del futuro, es muy usada en la venta cruzada, la detección de fraude, gestión de riesgo, etc. [15].

C. ANÁLISIS PRESCRIPTIVO

Analiza todos los elementos de una decisión conjuntamente con sus variables para establecer la probabilidad que cada uno se produzca, es posible tener mejor control de situaciones y aumenta la probabilidad que el resultado deseado suceda, se refiere a caracterizar los hechos con las diferentes variantes [10], es un sistema de aprendizaje no supervisado. Es bastante práctica, más conveniente y real ya que recomienda las mejores acciones basados en los análisis predictivos y descriptivos. [11]

D. ANÁLISIS CAUSAL

Busca los motivos de los resultados de los objetivos de la empresa propuestos y encontrar las causas que originan los problemas para proponer una solución en los diferentes niveles de una empresa [12]

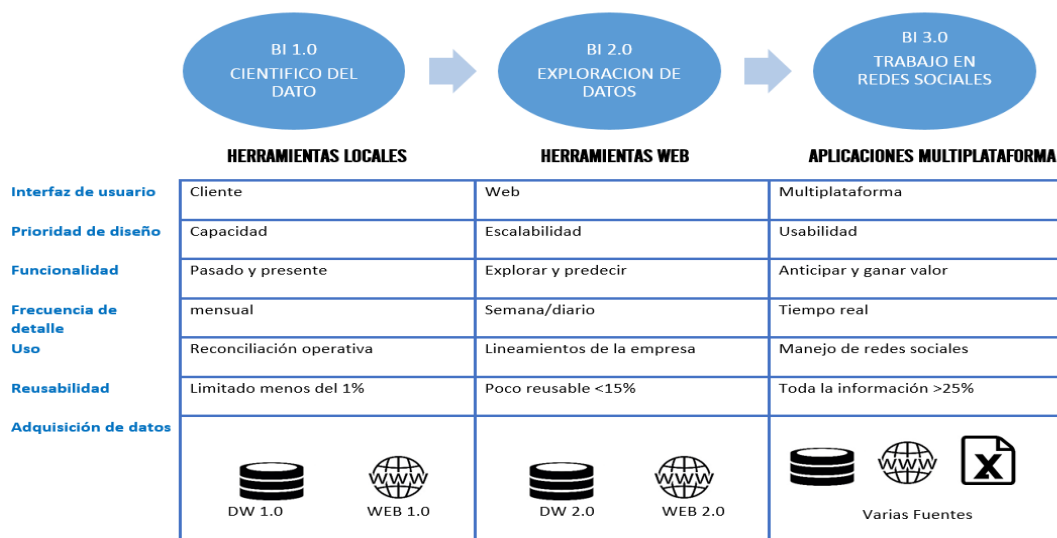


Figura 1. Evolución de BI con las características de cada versión y sus funcionalidades con diferentes accesos a datos.

IV. SERVICIOS Y APLICACIONES

Una plataforma de BI su principal función es proporcionar reportes a diferentes niveles de la empresa para la toma de decisiones, esta visualización debe ser interactiva y transparente de las consultas realizadas a las bases de información.

A. CLOUD BI

El almacenamiento de big data requiere una gran infraestructura de TI, los servicios en la nube (Cloud) permite reducir costes, eliminar licencias siendo flexible y disponible. Con servicios de las arquitecturas de red en la nube con capas de gestión, seguridad y desarrollo de aplicaciones, cada empresa debe decidir el nivel de uso del cloud para el desarrollo de BI entre: SaaS (Software como Servicio), PaaS (Plataforma como Servicio) y IaaS (Infraestructura como servicio). Las mayores empresas que proporcionan estos servicios es Amazon con su plataforma AWS (Amazon Web services), Verizon, BlueLock, CSC, Gogrid, IBM, OpenStack, Terremark [13].

La implementación en la nube de BI se necesita una estrategia bien definida orientada a servicios a nivel de la organización, una estrategia de Cloud BI contiene las siguientes 6 etapas. La iniciación para identificar la estrategia o el objetivo de negocio, en análisis y evaluación de estado del negocio, diseñar la solución de Cloud BI, la evaluación del proveedor de Cloud, la implementación de la estrategia de Cloud BI y finalmente la gestión del Cloud [14].

B. SOCIAL BI

La tendencia a redes sociales permite a BI interactuar de mejor manera con los usuarios mediante un análisis profundo de información para ofrecer nuevas oportunidades de conocer deseos, necesidades y demandas de un cliente. Social BI se presenta para sistematizar los datos de las redes sociales y generar modelos de datos multidimensionales para la demanda de productos, social BI es conocido también como CRM social. [15]

C. REAL-TIME BI

Se trata de reducir la brecha entre los datos al ingresar al sistema y cuando ellos pueden ser procesados para conseguir información para la toma de decisiones a su brevedad, REAL-TIME BI no tiene solo como objetivo procesar información lo más rápido posible, su objetivo es procesar en el tiempo indicado el volumen de datos necesarios, este proceso permite adaptarse con mayor rapidez a las nuevas tendencias de negocio [16].

D. MOBILE BI

El 33 % de los reportes de dashboard (tablero de instrumentos) se lo realizan desde los celulares y el 60% de corporaciones usan algún tipo de dispositivo móvil para los reportes [4], son los recursos físicos para el acceso de todos los servicios de BI, ya que puede monitorear en tiempo real los KPI's (métricas del negocio) de la empresa, pueden los empleados entregar información desde cualquier lugar, se generan alertas en ventas y brinda soporte a decisiones urgentes. Todas las empresas dedicadas a BI están migrando al servicio Mobile BI por una necesidad del mercado, su reto más importante es complementar las limitaciones físicas de

los equipos como el tamaño de la pantalla, la memoria y el consumo de batería [17].

E. BUSINESS ANALYTICS

Es el análisis del rendimiento de un BI, es la etapa que no tiene que ver con la infraestructura, hace relación a la efectividad de las tomas de decisión, es encuentra en la etapa entre el Cloud BI y la estadística aplicada, es el enfoque de negocios de un BI, software como IBM Cognos [18] explota las herramientas de BI y brinda formas de mediad de deciciones en relación a KPI's de la empresa, Cincinnati Zoo y Amarak son empresas que se enfocan al desarrolo de un sistema de medición de toma de decisiones y la correcta visualizacion de información.

V. MACHINE LEARNING EN BI

Machine learning proporciona métodos para tratar y extraer información de un conjunto de datos de forma automática, donde las personas no lo pueden realizar por el volumen y la complejidad de la información, se pueden clasificar a machine learning en dos grupos, supervisado y no supervisado [19], en técnicas supervisadas el objetivo es crear modelos a base de una entrada de datos y generar una salida, es decir conocer hacia donde se quiere llegar $Y=f(x)$, se pueden agrupar en clasificaciones o categorías y las regresiones que pronostican en el tiempo. Las técnicas no supervisadas son donde sólo se tiene datos de entrada y no existen variables correspondientes, el fin es aprender nuevas formas de correlación entre los datos para presentar nuevas estructuras, se pueden dividir en problemas de agrupación de los datos o la asociación entre ellos, además existen algoritmos híbridos que pueden ser supervisados y no supervisados como las redes neuronales [20], se detallan los algoritmos más usados en relación a las clasificación antes expuesta.

A. ALGORITMOS DE PREDICCIÓN

Permiten estimar un evento en relación a los datos de entrenamiento que tenga el modelo, son algoritmos supervisados.

- Regresión lineal:

Permite determinar el grado de dependencia de las series de valores X e Y, prediciendo el valor y estimado que se obtendría para un valor x que no esté en la distribución. [21] [22].

- Regresión exponencial

Es un modelo aplicado cuando la regresión lineal no puede determinar un coeficiente de determinación apropiados y se establece la tendencia a través de un diagrama de dispersión [21] [22].

- Estadística bayesiana

Enfoque de inferencia estadística mediante el uso de probabilidades bayesianas en las pruebas del modelo y experimentos [23].

B. ALGORITMOS DE CLASIFICACIÓN

Son algoritmos que buscan correlaciones entre los datos para agruparlos en relación a una clase o variable.

- Árbol de decisión

Son clasificadores que aproximan una función a partir de la ejecución de un conjunto de pruebas sobre los valores asociados a atributos predefinidos [24] [25] [26].

- KNN

El algoritmo asigna la clase a un nuevo ejemplo basado en las observaciones de las clases de sus vecinos más cercanos en relación a una distancia [24] [25] [26].

- LVQ

Se divide el espacio en regiones caracterizada por un vector con n-dimensiones [24] [25] [26]

C. ALGORITMOS DE CLÚSTER

Son algoritmos no supervisados que busca correlaciones entre los datos para agruparlos.

- K-means

Construye una solución aproximada al problema de minimización de distancia inter-grupo, escogiendo k centroides de forma que cada elemento se asigne al grupo relacionado al centroide más cercano [24] [25] [26].

- K-medians

Es una variación de K-means que n para determinar su centroide usa la mediana [24] [25] [26]

- Expectation Maximization

Se utiliza para encontrar estimadores de máxima verosimilitud en modelos de variables no observables [24] [25] [26]

D. REDES NEURONALES

Tratan de imitar el funcionamiento de una neouronas de organismos vivos, es decir trabajar en conjunto sin que exista una tarea concreta, con el tiempo generan conexiones entre si para aprender de casos anteriores.

- Back-propagation

Algoritmo para entrenar neuronas mediante un ciclo de propagación entre capas neuronales hasta generar una salida [24] [25] [26]. Arquitecturas

VI. DISCUSIÓN Y COMENTARIOS FINALES

BI es la plataforma de una empresa para el análisis de información para la predicción de patrones del mercado y se pueda tomar una decisión correcta mediante la visualización interactiva de información. La infraestructura BI con sus múltiples servicios en la nube permite dar un correcto rumbo de negocio a las empresas que se basan en algoritmos de aprendizaje de máquina para realizar diferentes tipos de análisis.

Existen muchas empresas encargadas de brindar la plataforma o el servicio de BI, pero es necesario contar con una metodología adecuada para la selección de la misma. El mayor reto de BI es el exponencial crecimiento de información que debe ser procesada en tiempo real con una visualización móvil.

VII. BIBLIOGRAFÍA

- [1] S. Bouchana y M. Janati Idrissi, «Towards an assessment model of end user satisfaction and data quality in Business Intelligence Systems,» *10th International Conference on Intelligent Systems*, n° 10, pp. 1-6, 2015.
- [2] R. Hans, «Modeling software engineering projects as a business: A business intelligence perspective,» *AFRICON*, pp. 1-5, 2013.
- [3] B. Nustara y Y. Lisanti, «Business Intelligence desing on the company,» *International conference on e-Education, Entertainment and e-Management*, pp. 356-361, 2011.
- [4] Gartner.com, «Magic Quadrant for Business Intelligence and Analytics Plataforms,» 04 Febrero 2016. [En línea]. Available: <http://gartner.com/>. [Último acceso: 29 Julio 2016].
- [5] L. James, «Yellowfin,» 10 Abril 2014. [En línea]. Available: <http://www.yellowfinbi.com/YFCommunityNews-Defining-Business-Intelligence-3-0-159445>. [Último acceso: 29 Julio 2016].
- [6] H. Chen, R. Chiang y V. Storey, «BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT,» *MIS Quarterly*, vol. 36, n° 4, pp. 1165-1188, 2012.
- [7] A. McAfee y E. Brynjolfsson, «Big Data: The Managment Revolution,» *Hardvard Business Review*, pp. 59-63, 2012.
- [8] S. Raza Quresh y A. Gupta, «Towards Efficient Big Data and Data Analytics: A review,» *IT in Business,Industry and Government (CSIBIG), 2014 Conference on*, pp. 1-6, 2014.
- [9] M. Mircea y M. Stoica, «Combining Business Intelligence with Cloud Computing to Delivery Agility in Actual Economy,» *CiteSeerX*, pp. 1-16, 2016.
- [10] S. Ouf y M. Nasr, «The Cloud Computing: The Future of the BI in the Cloud,» *PorQuest*, pp. 750-761, 2011.
- [11] M. Rosemann y M. Eggert, «LEVERAGING SOCIAL NETWORK DATA FOR ANALYTICAL CRM STRATEGIES - THE INTRODUCTION OF SOCIAL BI,» *AIS Electronic Library*, pp. 95-101, 2012.
- [12] G. Finnie y J. Barker, «Real-Time Business Intelligence in Multi-Agent Adaptive Supply Network,» *ePublications*, pp. 1-5, 2005.
- [13] I. Chan, H.-K. Tan, P. Tan Lau y W. Yeoh, «State of the Art Review and Critical Success Factors for Mobile Business Intelligence,» *IBIMA Publishing*, pp. 1-10, 2013.
- [14] J. Evans y C. Lindner, «Business Analytics: The Next Frontier for Decision Sciences,» *Desicion Science Institute*, 2002.
- [15] L. Solutions, Predictive Analytics, lantares.com, 2012.
- [16] D. Bell y H. T. A. Raiffa, *Decision making Descriptive, normative and prescriptive interactions*, Cambridge: Cambrigde University Press, 1999.
- [17] F. S. Pares, «DBA Business Analytics,» DBA, 2015. [En línea]. Available: <http://francescsanchezbi.webnode.es/news/estado-actual-del-analisis-prescriptivo/>. [Último acceso: 15 Agosto 2016].
- [18] foromarketing, «FMK,» [En línea]. Available: <http://www.foromarketing.com/diccionario/analisis-causal/>. [Último acceso: 15 Agosto 2016].
- [19] J. Berral-García, «A Quick View on Current Techniques and Machine Learning Algorithms for Big Data Analytics,» *ICTON*, pp. 1-4, 2016.
- [20] J. Brownlee, *Machine Learning Algorithms, Machine Learning Mastery*, 2015.