Journal Business Science vol. 1 No. 2 (2020) 94 - 104

JBS JOURNAL BUSINESS SCIENCE

Journal Business Science

https://revistas.uleam.edu.ec/index.php/business_science

ISSN impreso 2737 - 6184 / ISSN electrónico 2737 - 615X

METODOLOGÍA M3S PARA PROYECTOS DE INTELIGENCIA DE NEGOCIOS

M3S METHODOLOGY FOR BUSINESS INTELLIGENCE PROJECTS

Santiago Morales Cardoso[1]; Glenda Toala Sánchez[2]

Universidad Central del Ecuador[1,2], Quito-Ecuador

smorales@uce.edu.ec[1]; gtoala@uce.edu.ec[2]

Santiago Morales Cardoso[1]https://orcid.org/0000-0002-3833-9654
Glenda Toala Sánchez[2]https://orcid.org/0000-0002-4596-3770

Recibido: 16/12/2020 **Aceptado: 25/12/2020**

Código Clasificación JEL: A12, C80, D70.

RESUMEN

En la actualidad es importante orientar a las organizaciones en una oportuna toma de decisiones, apoyadas por todas las ventajas que proporciona la Inteligencia de Negocios (IN), la Metodología M3S es una alternativa innovadora, con grandes aportes en los procesos de la transformación de los datos con la finalidad de mejorar la toma de decisiones, esta propuesta metodológica incluye un algoritmo de decisión, aportando con mejoras en la búsqueda de sus nodos por medio de atributos ubicados en diferentes fuentes de datos estructurados o no, los mismos que pasan por procesos de limpieza especial y a los cuales se les da en este estudio una denominación de “puros”. Así, obtenidos los nodos, estos pasan a ser clasificados en base a los cálculos recursivos de entropía y ganancia de información para lograr una estructura robusta que tome decisiones para cualquier dato que ingrese en un determinado caso de estudio.

Palabras Clave: Entropía, Clasificación, Limpieza de Datos, IN, Árboles de decisión.

ABSTRACT

Nowadays it is important to guide organizations in an opportune decision making, supported by all the advantages that Business Intelligence (BI) provides. The M3S Methodology is an innovative alternative, with great contributions in the processes of data transformation with the purpose of improving decision making, This methodological proposal includes a decision algorithm, contributing with improvements in the search of its nodes by means of attributes located in different sources of structured or unstructured data, the same ones that go through special cleaning processes and which are given in this study a denomination of "pure". Thus, once the nodes are obtained, they are classified based on the recursive calculations of entropy and information gain to achieve a robust structure that makes decisions for any data entered in a given case study.

Keywords: Entropy, Classification, Data Cleaning, BI, Decision Trees.

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

INTRODUCCIÓN

"La inteligencia de negocios, es un enfoque estratégico para orientar sistemáticamente el seguimiento, la comunicación y la transformación relacionada al débil conocimiento de la información procesable en la cual se basa la toma de decisiones" (Kamel y Samia, 2002). En la época pasada los Data Warehouse (DW) han sido ampliamente adoptados en los negocios, proveían análisis multidimensional con muchos datos históricos del negocio, ayudando a crear nuevas decisiones. "Sin embargo, en el 20% de las empresas se encuentran los datos numéricos y el otro 80% están escondidos en NO numéricos y hay que recuperarlos para realizar un análisis más completo de Business Intelligence" (Tseng & H,, 2006).

"Las recomendaciones para implementar un proyecto de Inteligencia de Negocios basado en un DW, debe analizarse desde diferentes aspectos: Económico (Costos cuantificables e intangibles), Social (Ambiente organizacional, Apoyo, Aspectos Administrativos, Compromiso), y Técnico (Infraestructura, Metodologías)". (Villalon, 2006). Cada organización maneja la Inteligencia de Negocios de acuerdo a las exigencias de la misma. Es aquí, que se ve la necesidad de crear una metodología en cuanto al proceso ETC para empezar a definir estándares, ya que los softwares que existen actualmente en el mercado no necesariamente ofrecen las mejores alternativas. "En el año 2003 ya se decía que la demanda de Sistemas de Inteligencia de Negocios estaba creciendo, pero la producción de software en general caminaba más lento" (Negash, 2004). De esta manera otro autor considera que "la investigación y la gestión del conocimiento es una idea básica, por lo que el conocimiento que se genere esté compartido a la comunidad" (Krogh & von, 2002).

El crecimiento en los servicios de consultoría de Reingeniería de Procesos de Negocio (BPR) ha llevado a una proliferación de métodos para conducir BPR. Se han hecho investigaciones revisando Métodos, Técnicas y Herramientas (MTTs), y en base a encuestas exhaustivas de estas técnicas de uso común de BPR, se forma una base de conocimientos para mejorar la práctica de cambio de procesos empresariales y proporciona una base para la investigación BPR futuro (Kettinger, Teng, & Subashish, 1997).

"Tomar mejores decisiones más rápido, convertir datos en información, y usar nuevo conocimiento para las empresas es actualmente una ventaja competitiva" (Calzada & Abreu, 2009). "La Inteligencia de Negocios se puede definir como el cúmulo de modelos matemáticos y metodologías de análisis que explotan los datos disponibles para generar información y conocimientos útiles para los complejos procesos de toma de decisiones" (García, 2012).

Debido a la importancia de esta definición en el estudio se consideran algunos autores más que indican: "Inteligencia de Negocio es un concepto que "integra como solución el almacenamiento y procesamiento de enormes cantidades de datos e información para transformarla en conocimiento y decisiones en tiempo real a través de una fácil explotación" (Salinas La Rosa, 2010); también indica que: "BI es un término que engloba las aplicaciones, infraestructura, plataformas, herramientas y mejores prácticas que permiten acceder y analizar la información para optimizar las decisiones y gestionar el rendimiento empresarial".

Uno de los criterios es que: "Se entiende por Inteligencia de Negocios al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización" (Díaz, 2012). Desde otro punto de vista se considera la siguiente definición, que asevera: "La inteligencia empresarial es una categoría de aplicaciones y tecnologías para recopilar, proporcionar acceso y analizar datos con el fin de ayudar a los usuarios empresariales a tomar mejores decisiones" (Ranjan, 2009).

Por último se revisa que la inteligencia de negocios se define como "la habilidad corporativa para tomar decisiones; esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos, y aplicar en ellos técnicas analíticas de extracción de conocimiento" (Parr & Sons, 2009), "los datos pueden ser estructurados para que indiquen las características del área de interés" (Stackowiak, Greenwald, & Rayman, 2007), "generando el conocimiento sobre los problemas y oportunidades del negocio para que pueden ser corregidos y aprovechados respectivamente" (Ballard, 2006) y Grossman (1998), mencionan que IN, "es un término general para un conjunto de conceptos y métodos para mejorar la toma de decisiones empresariales, utilizando sistemas basados en hechos".

Actualmente, se ha detectado que uno de los retrasos en proyectos de IN se produce en los procesos de Extracción, Transformación y Carga de Datos (ETC), para lo cual se propone una metodología que

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

optimice estos procesos con la ayuda de árboles de decisión que aplican aprendizaje automático en las decisiones.

En los proyectos de analítica de datos es necesario permanentemente buscar un nuevo modelo que guíe a conseguir soluciones en el ámbito de inteligencia de negocios ya sea considerando soluciones tradicionales o aquellas de las que actualmente se las conoce como In Memory. Se propone una metodología que explote y mejore principalmente el proceso extracción y transformación de datos, pero introduciendo aprendizaje automático a través del uso de algoritmos de árboles de decisión que dentro del estudio se les dará mayores parámetros de valor por lo que se considerará como ideales. Esto ayudará a tener soluciones que beneficien la creciente necesidad que tienen los ejecutivos y en general los usuarios de diferentes niveles jerárquicos en las empresas de contar con información tomada de distintas aristas pero que lleguen limpia, que les permita mejorar sus decisiones en base al mayor conocimiento que se llega a tener de las organizaciones. Tomando en cuenta que en la mayoría de los casos la información se encuentra dispersa en varios sistemas y en diferentes fuentes de datos, esto implica que debe crearse una metodología que logre consolidar la información en base a lo aprendido y en el presente y futuro pueda colaborar con la toma de decisiones certeras en las empresas.

La Metodología M3S que incluye un algoritmo de decisión basado en ID3, fundamenta, representa y gestiona la implementación de proyectos de IN usando modelos matemáticos y explotando datos disponibles para generar conocimiento tomando en cuenta el tiempo del proyecto, madurez de la información, introducción de técnicas de Inteligencia Artificial como los árboles de decisión y aprendizaje automático.

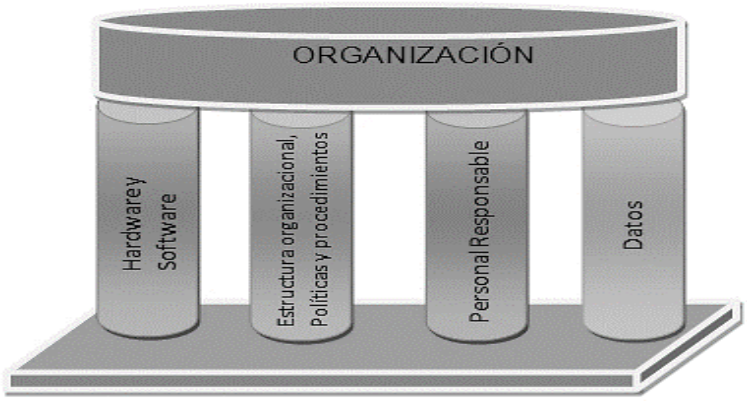
M3S, al recopilar datos de diferentes fuentes heterogéneas, se considera que es más robusta y puede tener mayor certeza luego de su procesamiento de pureza de información, a diferencia de datos planos, que generan variedad de escalas de tiempo, las mismas que determinan los parámetros de integración o identifican los acoplamientos entre estas series, como bien se lo menciona en la Entropía de Permutación.

Al aplicar la Metodología M3S en casos de estudios específicos, se observó una ganancia competitiva tanto en la velocidad de obtención de los datos como en su calidad, que permitió ampliar la visión estratégica del negocio, reducir riesgos de costos y disminuir la incertidumbre en la toma de decisiones.

METODOLOGÍA

La metodología de implementación para procesos de IN, cuenta con las técnicas necesarias para el seguimiento de proyectos, así como con una estructura organizacional de los mismos, considerando cuatro factores críticos de éxito, los cuales deben ser base para cualquier tipo de proyecto de tecnología (Ver Figura 1), que permitirá conseguir una buena rigurosidad en su desarrollo, de igual forma permite un control adecuado en todos los procesos y actividades que se vayan ejecutando.

Figura1: Pilares de la estructura organizacional



Fuente: Morales, 2019

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

Para la elaboración de la propuesta metodológica se utilizaron los instrumentos de encuestas, entrevistas y casos de estudio. En el componente cualitativo, se midieron variables como la satisfacción, madurez de la información y calidad del entorno. Para el componente cuantitativo se establecieron variables para determinar el nivel de mejoramiento en tiempos de extracción y transformación de datos, valorados en formularios creados para esa finalidad.

La hipótesis formulada es la siguiente: La metodología M3S integra y optimiza las propuestas actuales en la extracción y transformación de datos dentro de un proyecto de IN.

Para la elaboración de la propuesta de la Metodología M3S (Morales, 2019), en primer lugar, se definieron las etapas del ciclo de vida, luego se analizó la etapa de diseño, es allí donde se propone una mejora en relación a las metodologías ya existentes, aplicando un modelo matemático basado en algoritmos de árboles de decisión.

PROPUESTA DE LA METODOLOGÍA M3S

El Ciclo de Vida que se definió en M3S está compuesto de seis etapas que se encuentran definidas de manera literal de la letra A hasta la F y dentro de la cual será en la etapa D de Diseño, en la cual se hizo énfasis y en la que se establece una nueva metodología dentro de M3S, que se denomina "Metodología de Decisión", basada en algoritmos de árboles de decisión conocidos en el estudio como "ideales", explicados en los siguientes puntos.

Las etapas del ciclo de vida de la Metodología M3S se detallan en la Tabla 1.

Tabla 1.
Ciclo de Vida de Metodología M3S

ETAPAS DEL CICLO DE VIDA DE LA METODOLOGÍA M3S PARA PROYECTOS DE INTELIGENCIA DE NEGOCIOS	
A. Justificación	A.1. Evaluación caso de negocio A.2. Acta de Constitución del Proyecto.
B. Planificación	B.1. Evaluación Infraestructura ▪ Técnica ▪ No Técnica B.2. Definición del Alcance B.3. Cronograma
C. Análisis y requisitos de negocio	C.1. Definición de requisitos de Proyecto C.2. Prototipo de Aplicación
D. Diseño	**D.1. Diseño ETC** **D.1.1. Diseño Metodología de Decisión**
E. Construcción	E.1. Desarrollo ETC E.1.1. Obtención de Metadata
F. Implementación	F.1. Puesta en marcha. F.2. Evaluación de la Solución F.3. Cierre del Proyecto

Fuente: Morales, 2019

En la primera etapa (A) de Justificación, se buscan los argumentos reales que generen la necesidad del negocio para implementar un proyecto de ingeniería dentro del cual se debe evaluar el caso de negocio en donde principalmente se debe justificar la inversión mostrando el balance entre el costo, tiempo de ejecución y los beneficios conseguidos; y también se debe generar el Acta de Constitución, que es un documento inicial que formaliza la ejecución del proyecto.

La siguiente etapa (B) es la de Planificación, que define la estrategia y planes tácticos a seguir para desarrollar y llevar a cabo el proyecto con éxito, en la que se debe evaluar la infraestructura técnica, así

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

como revisar las políticas institucionales, personal que se asigna al proyecto y logística de oficina. Es importante definir el alcance del proyecto, incluyendo criterios específicos, medibles, realistas y enmarcados en un tiempo establecido, llegando a definirse claramente el objetivo del análisis de IN; y por último establecer un cronograma, en donde se define cada actividad designando los responsables, fechas de hitos importantes a cumplirse y a controlarse.

Continuando con la etapa (C), Análisis y Requisitos de Negocio, en la cual se realiza un análisis detallado del problema y oportunidades del negocio, para conseguir una comprensión sólida de los requisitos de la organización que sean necesarios para llegar a una posible solución o producto. Se debe dentro de lo posible proponer un prototipo de lo que llegaría a presentarse en los resultados de tableros de medición.

En la cuarta etapa (D) que es el Diseño, concibe una alternativa que resuelva el problema para la toma de decisiones, dentro de la cual se tiene el Diseño ETC en donde los datos de origen para las aplicaciones de IN provienen de varias plataformas heterogéneas, que son gestionadas por una variedad de sistemas operacionales y aplicaciones. Partiendo de la premisa de que los sistemas deben aprender y adaptarse a su entorno, el propósito de los procesos ETC en esta investigación es unir los datos de estas plataformas y transformarlos a un formato que contenga todas las clasificaciones de información necesarias, de tal forma que a partir de estos se tenga capacidad de decidir sobre todo el universo de información actual y la futura que llegue para un determinado objetivo. Por ser una parte esencial de la investigación esta parte tendrá un tratamiento especial luego de la explicación de las dos últimas etapas de M3S.

En M3S la etapa (E), es la Construcción, en la cual se aplica la metodología de decisión en un software que realice todo lo diseñado en la etapa anterior (D), dentro de la que se ingresará en un proceso de software las bases necesarias, las reglas definidas a fin de que la tabla de datos pueda ser analizada y poder inferir el árbol de decisión ideal.

La última etapa (F) es la de Implementación, conocida también como venta del producto acabado, y entonces medir su efectividad para determinar si la solución encontrada, excede o no el resultado esperado. Se pone en marcha de forma tradicional, o incremental según el mejor criterio de los expertos de BI. De igual forma se evalúa la solución. A esta altura lo primordial es cerrar el proyecto, en donde se debe archivar toda la documentación generada para que se encuentre disponible a cualquiera de los niveles empresariales, formalizando la entrega de todos los aplicativos y aceptación del cliente en base a firmar el acta de entrega total del proyecto.

Como se indicó anteriormente la propuesta se concentra en la etapa de Diseño de M3S, estableciendo una metodología de decisión y las fases que la componen.

Metodología de Decisión de la Etapa de Diseño

Antes de presentar la metodología de decisión, se debe realizar ciertas consideraciones como expresar que la extracción es independiente del origen de datos que se tenga, los mismos que pueden ser cualquier motor de base de datos relacionales, un Datawarehouse, archivos planos, hojas Excel, redes sociales entre muchos otros. “Los árboles de decisión se utilizan desde los años cuarenta y son especialmente apropiados para expresar procedimientos médicos, legales, comerciales, estratégicos, matemáticos, lógicos, entre otros, estos se caracterizan por la sencillez de su representación y de su forma de actuar, además de la fácil interpretación, dado que pueden ser expresados en forma de reglas de decisión” (Martínez & R., 2009); el centro metodológico se basará en los árboles de decisión ideales, para los que previamente se debe establecer reglas de negocio claras, las mismas que lleven a generar una tabla temporal en donde van a ser ubicados los mejores atributos en su categorización establecida y en el número de ellos que sean necesarios, los cuales agregarán mayor y mejor información para la toma de decisiones y en muchos casos para predicciones futuras.

Fases de la Metodología de Decisión

Es importante definir un proceso de limpieza de datos como el “proceso de asegurar que todos los valores en un conjunto de datos sean consistentes y correctamente registrados” (Urdaneta, 2001); Galhardas señala que el problema consiste en: “remover inconsistencias y errores desde los conjuntos originales de datos, es bien conocida en el área de Sistemas de Soporte de Decisiones y DW” (Galhardas, Florescu, & Simón, 2000).

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

Se debe llegar a que los datos con niveles de pureza que maneja M3S sean consistentes y con alta credibilidad, por lo que se propone el siguiente flujo de limpieza:

1. Validación de valores nulos o vacíos
2. Tratamiento de registros duplicados.
3. Filtrar los datos semejantes.
4. Homogeneizar los datos de bases heterogéneas que vinieron diferentes orígenes estructurados o definitivamente no estructurado como redes sociales.

Y con estas consideraciones especiales de la conformación de los elementos del árbol de decisión ideal, ahora se analiza lo descrito anteriormente sobre el Diseño ETC, poniendo las siguientes fases que son descritas con cada una de sus actividades en la Tabla 2, para la aplicación del Algoritmo Ideal dentro de la Metodología de Decisión que será implementada en M3S:

Tabla 2.
Etapas de Metodología de Decisión

FASES	ACTIVIDADES
1. Análisis de Calidad de Datos	Analizar datos para un proceso de decisiones, mediante uso de diccionarios de datos, estadísticas, redes sociales y conocimiento de expertos; escogiendo los atributos más relevantes para el estudio, con las características de pureza explicadas anteriormente, sean del tipo 1 (estructurado) o tipo 2 (no estructurado).
2. Hipótesis de Datos	El objetivo que tendrá esta parte metodológica es justamente encontrar una descripción del problema planteado, buscando mediante suposiciones generales a partir de hechos concretos y a raíz de estos clasificar la información para generalizar los casos nuevos.
3. Generación de la tabla para Entrenamiento	Esta tabla es el producto final de los datos limpios y lo más refinados posible según su categoría; así también de la formalización de la o las reglas de negocio previamente establecidas con las cuales se han logrado clasificar la información. A raíz de esta tabla se puede generar el proceso matemático y estadístico para la creación del mejor árbol de decisión de acuerdo con el algoritmo ideal propuesto y a la objetividad que se haya dado en el análisis con la información encontrada.
4. Inferencia del Árbol de decisión	Para generar el árbol de decisión se utilizará inferencia inductiva que es aprendizaje supervisado. Este árbol es una forma de la aplicación del algoritmo ID3 con las mejoras específicas sobre la composición de nodos de dos tipos, el cual se basa en las reglas del negocio formalmente elegido; y del que se prevé tener la mayor ganancia de información, el mismo que se encuentra detallado en la etapa de construcción. Se procederá de forma recursiva de arriba hacia abajo; en cada nodo el atributo "más importante" que discrimina los ejemplos que han llegado hasta este nodo y por último clasificando los casos que siguen para el siguiente nivel.
5. Pruebas de datos sobre el árbol de decisión	Una vez que se cuenta con el árbol ideal de decisión, el cual con los cálculos de entropía genera la mayor ganancia de información; se elabora un test para verificar que siempre se obtendrán los mejores resultados con nuevos datos que ingresen a evaluarse en el árbol. Se debe tomar en cuenta que, para algunos casos al probar nuevos datos, estos podrían hacer que cambie la clasificación y reglas de la tabla de entrenamiento, lo cual produciría una alteración en el árbol de decisión generado.

Fuente: Morales, 2019

A continuación, se representa en forma de un modelo matemático cada uno de los pasos contemplados en la metodología de decisión, representado en base a funciones los datos y a su vez englobándolos como conjuntos o grupos bajo su pertenencia. Se usan fórmulas como la de entropía y ganancia de Información que definen la posición de cada Nodo (Atributo) dentro del árbol de decisión ideal que se llegó a obtener:

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

1. **Análisis de Calidad de Datos**
   Dado A conjunto de datos estructurados, y
   Dado B conjunto de datos No estructurados ponderados (1.5)

Si X ∈ A ó X∈ B

Sea $f: X -> Y$, en donde:
X son todos los casos posibles de entradas
Y espacio de salidas

2. **Hipótesis de Datos (H)**
   Serán todas las $h \in H$

3. **Generación de la tabla para Entrenamiento:**
   Si $f \in H$ entonces es Aprendizaje considerado
   Si $f \notin H$ entonces no será considerado.

4. **Inferencia de Árbol de decisión**
   Usar conocimiento a priori para asegurar que $H \subset f$ en donde la inferencia se lo realiza mediante:
   - Cálculo de la entropía

   Dados:
   - Un problema con dos clases (positiva y negativa)
   - S, el conjunto de ejemplos.

$$Entropia(S) = \sum_{i}^{k} p_i \log_2\left(\frac{1}{p_i}\right) \; con\; k\; clases$$

   - Y también mediante la ganancia que se obtendrá

$$Ganancia(S, A) = Entropia(S) - \sum_{v\, e\, valores\, de\, A} \frac{|S_v|}{|S|} Entropia(S_v)$$

5. **Pruebas de datos sobre el árbol de decisión:**

Se utilizará medidas comunes de evaluación que son:

$$promedio\ positivo\ verdadero = \frac{TP}{TP + FN}$$

$$promedio\ positivo\ falso = \frac{FP}{FP + TN}$$

$$precisión = \frac{TP}{TP + FP}$$

$$correcciòn = \frac{TP + TN}{TP + FP + TN + FN}$$

Dónde: TP: Verdaderos positivos (clasificado: positivo, realidad: positivo)
TN: Verdaderos negativos (clasificado: negativo, realidad: negativo)
FP: falsos positivos (clasificado: positivo, realidad: negativo)
FN: falsos negativos (clasificado: negativo, realidad: positivo)

Se considera que "Un árbol de decisión es un conjunto de datos donde se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema" (Rokach & Maimon, 2008).

Hay que tomar en cuenta que los mejores atributos para toma de decisiones es mediante la generación de aprendizaje automático; se dice, también que "es un aspecto crucial del comportamiento inteligente de los seres humanos, que les permite acumular experiencias y adaptarse a entornos cambiantes frente a los

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

que deben utilizarse nuevas estrategias" (Pino, Gómez, & de Abajo, 2001), otro autor dice que "es desarrollar técnicas que permitan a las computadoras aprender, crear programas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos, induciendo conocimiento" (Moreno, 1994).

Se revisará que aprendizaje supervisado es el algoritmo que produce una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema, como lo es el problema de clasificación, donde el sistema de aprendizaje trata de etiquetar (clasificar) una serie de vectores utilizando una entre varias categorías (clases) (Salas, 2004), como lo es el Modelo de clasificación ID3, que significa "inducción mediante árboles de decisión" que aplica la estrategia "divide y vencerás" para hacer la clasificación, implementando métodos y técnicas para la realización de procesos inteligentes, representando así el conocimiento y el aprendizaje, automatizando tareas.

Recopilando nuevas definiciones de Entropía, la cual toma en cuenta la incertidumbre, impureza, desorden de un conjunto de datos relativo a una clasificación binaria, a su vez se define matemáticamente como la propiedad de que si se asigna p = 1 a una de las respuestas, consecuentemente p = 0 a todas las demás, en cuyo caso la entropía resulta ser cero. Y si todas las probabilidades son iguales, la entropía es máxima (García, 2018); y por último se recuerda a la Ganancia de Información, la cual ocurre por el menor valor de entropía esperado (S), al calcular todo el subconjunto de atributos que fueron evaluados según la calidad de información que guardan.

De todo el detalle que se encontró en la revisión de M3S, se puede concluir en primer lugar que las eficiencias de los algoritmos de árboles de decisión dependen en gran medida de la pureza de información que se pueda conseguir, el conocimiento que entreguen los expertos de negocio, para generar una óptima tabla de entrenamiento (TE), sobre la cual la implementación de ID3-Modificado como se lo ha llamado, entrega un árbol de decisión ideal que mantiene la mayor ganancia de información posible, que, sin embargo, está sujeto a cambios en base a nueva información que puede ingresar al sistema. También es válido indicar que el definir una tabla de entrenamiento que mantiene todas las agrupaciones concretas de datos, garantiza la fiabilidad de las decisiones inducidas, por eso es por lo que, al momento como resultados iniciales, se creó una Metodología que aporta como guía dentro de ETC, en la cual lo más relevante es que se complementa un algoritmo de aprendizaje para extraer información valiosa de las empresas.

DISCUSIÓN

Para el desarrollo de proyectos de IN, existen varias metodologías genéricas (Cruz Caicedo, 2012), como ISO21500, PMBOK y PRINCE2, las mismas que son compatibles entre sí y pueden ser complementarias según el tipo de proyecto, por lo tanto, para elaborar la propuesta de M3S se consideran algunos aspectos positivos de cada una de ellas, en el caso de PMBOX, destaca su eficiencia en los procesos y gestión de tiempos, PRINCE2, tiene su fortaleza en la gestión de cambio. Sin embargo, se identificaron los aspectos débiles, tal como los problemas generados al aplicar ISO21500 en el control de los proyectos en cuanto a las tareas en el ciclo de vida, al no considerar su tamaño, así como en la gestión de cambios.

De igual manera se revisan los modelos INMON y KIMBALL, en donde KIMBALL muestra mayor facilidad en la orientación de los procesos y su modelado, además permite encontrar indicadores de gestión en forma más efectiva, mientras que INMON mantiene un enfoque diferente, visualizando a la organización en conjunto.

Una vez elaborada la propuesta de M3S, con la finalidad de medir su eficacia, se la aplicó en el desarrollo de proyectos de IN en diferentes organizaciones, obteniendo resultados relevantes en comparación con el uso de otras metodologías, como la mejora en tiempo y en el número de personas involucradas en el desarrollo de proyectos, lo cual implica una reducción de costos, generando de esta manera una mayor rentabilidad, logrando así optimizar los procesos de la transformación de datos para la toma de decisiones.

Considerando que la IN permite buscar una nueva forma de analizar, diseñar, construir y ejecutar proyectos de este tipo, la contribución principal de esta propuesta es brindar un nuevo conocimiento sobre la información propia de cada organización, con la finalidad de permitir la toma de decisiones de manera más eficiente, mediante el uso de la metodología M3S, cuyo aporte significativo es la implementación de

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

una metodología de decisión dentro de la etapa de diseño del ciclo de vida del proyecto, basada en algoritmo de árboles de decisión, que sirvió para la verificación de los datos y predecir las decisiones, además de disminuir el tiempo invertido, de esta manera facilita a las pequeñas y medianas empresas a gestionar sus procesos de innovación tecnológica.

CONCLUSIONES

En los casos que se han analizado se almacenaba enorme cantidad de datos con un alto grado de heterogeneidad, pero la clave radicó principalmente en que se ganó ventaja competitiva tanto en la velocidad de obtención como en la misma calidad y nivel de pureza de esos datos agrupándolos según la categoría aquí descrita.

- M3S, es una metodología que apunta a las nuevas posibilidades de IN, con fuerza en las pequeñas y medianas empresas públicas o privadas y de cualquier sector productivo o comercial, de tal forma que soportados en la innovación tecnológica se implantó para mejorar las decisiones en las empresas estudiadas.
- M3S se la ha socializado para empresas con un mínimo nivel de madurez en el manejo de la información.
- M3S se construyó en base a la experiencia pragmática de profesionales con experticia de muchos años en proyectos de este tipo, por lo que justamente pudo ser probado en diferentes empresas y con el apoyo necesario.
- Luego de los casos de estudio se considera que se implantó mejor en pequeñas y medianas empresas por la mayor flexibilidad en su estructura plana organizacional, lo que llevó a un mejor apoyo tanto técnico como administrativo.
- M3S, se ejecutó con una clara descripción de los procesos y actividades a seguir en la metodología, haciéndola mucho menos burocrática incluso en su propia documentación, más aún en sus etapas que se definen como más ágiles.
- Se presentó como una metodología con mayor flexibilidad en la medida que la experiencia de los profesionales que estén al frente puedan ejecutarla de manera óptima en cada una de sus etapas.

REFERENCIAS BIBLIOGRÁFICAS

Cruz Caicedo, L. (2012). *Comparativa ISO 21500 y PMBOK Versión 5*. I Congreso Internacional Gerencia de Proyectos. Bogotá, Colombia.
http://americalatina.pmi.org/media/files/latam/colombia/2012_co_cruz.aspx

Díaz, J. C. (2012). Introducción al Inteligencia de Negocios (Vol. 1era.). (E. C. SA, Ed.) UOC.

Galhardas, H., Florescu, D., Shasha, D., & Simon, E. (2000, May). AJAX: an extensible data cleaning tool. In ACM Sigmod Record, *29(2),* 590.

García Reyes, R. (2012). *Minería de Datos para la Toma de Decisiones e Inteligencia de Negocios. México D.*F., Universidad Nacional Autónoma de México.

García, R. (12 de febrero de 2018). Repositorio Central Cepal. http://repositorio.cepal.org/bitstream/handle/11362/33747/S8500576_es.pdf?sequence=1

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

Kamel, Rouibah y Samia, O. (2002). A concept and prototype for linking business intelligence to business strategy. Journal of Strategic Information System, *l(11),* 133-152. http://www.sciencedirect.com/science/article/pii/S0963868702000057

Krogh, G., & von. (2002). The Communal Resource and Information System. Journal of Strategic Information System, *11*, 85-107. http://www.sciencedirect.com/science/article/pii/S0963868702000069

Martínez, R. E., Ramírez, N. C., Mesa, H. G., Suárez, I. R., Trejo, M. D., León, P. P., & & Morales, S. L. (2009). Árboles de decisión como Herramienta en el Diagnóstico Médico. http://www.medigraphic.com/pdfs/veracruzana/muv-2009/muv092c.pdf

Morales Cardoso, S. L. (2019). *Metodología para procesos de Inteligencia de Negocios con mejoras en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones.*

Moreno, A. (1994). Aprendizaje Automático. Ediciones UPC.

Negash, S. (2004). Business Intelligence. Communications of the Association for Information Systems, *13,* 177-195. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3234&context=cais

Parr, O. R., & Sons. (2009). Business Intelligence Success Factors.

Pino, R., Gómez, A., & de Abajo, N. (2001). Aprendizaje automático. En Introducción a la inteligencia artificial (8). Servicios de Publicaciones Universidad de Oviedo.

Quinlan, J. (1986). Induction of decision trees. Machine learning. Kluwer Academic Publishers. https://doi.org/10.1007/BF00116251

Ranjan, J. (2009). Business Intelligence: Concepts, components, Techniques and benefits. Journal of Theorical and Applied Information Technology, *70*. http://www.jatit.org

Rokach, L., & Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Publishing Co. Pte. Ltd.

Salas, R. (2004). Redes Neuronales Artificiales. Departamento de Computación.

Salinas La Rosa, A. (2010). Inteligencia de Negocio. Auditoría y Control. Prototipo de Herramienta de Calidad (Vol. 1). (D. D. Informática, Ed.) Universidad Carlos III de Madrid. https://core.ac.uk/download/pdf/29402111.pdf

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)

Stackowiak, R., Greenwald, R., & Rayman, J. (2007). Oracle Data Warehousing and Business Intelligence

Tseng, F. S., & H, C. A. (2006). The concept of document warehousing for multi-dimensional modeling of textual-based business intelligence. Decision Support System, 727-744.

Urdaneta,Elymir (2001). *El Data Mining.*

Villalon M, D. M. (2006). Recomendaciones para la Implementación de BI, basado en Data Warehouse. Técnico de Monterrey. http://hdl,handle.net

Licencia de Creative Commons (http://creativecommons.org/licenses/by-nc-sa/4.0)