

Diseño de una herramienta de identificación de rutas de aprendizaje basado en patrones de consumo de Objetos de aprendizaje en la plataforma MOODLE

Olga Beatriz España Rios, ✉ obespana@hotmail.com

Monografía presentada para optar al título de Magíster en Ingeniería de Software

Asesor: Rocio Emilia Segovia Jimenez, Magíster (MSc) en Ingeniería



Universidad de San Buenaventura Colombia

Facultad de Ingeniería

Maestría en Ingeniería de Software

Santiago de Cali, Colombia

2018

Citar/How to cite	[1]	
<hr/>		
Referencia/Reference	[1]	O. B. España Ríos, “Diseño de una herramienta de identificación de rutas de aprendizaje basado en patrones de consumo de Objetos de aprendizaje en la plataforma MOODLE”, Trabajo de grado Maestría en Ingeniería de Software, Universidad de San Buenaventura Cali, Facultad de Ingeniería, 2018.
Estilo/Style:		
IEEE (2014)		



Maestría en Ingeniería de Software, Cohorte 3
LIDIS.

Bibliotecas Universidad de San Buenaventura



Biblioteca Digital (Repositorio)
<http://bibliotecadigital.usb.edu.co>

- Biblioteca Fray Alberto Montealegre OFM - Bogotá.
- Biblioteca Fray Arturo Calle Restrepo OFM - Medellín, Bello, Armenia, Ibagué.
- Departamento de Biblioteca - Cali.
- Biblioteca Central Fray Antonio de Marchena – Cartagena.

Universidad de San Buenaventura Colombia

Universidad de San Buenaventura Colombia - <http://www.usb.edu.co/>

Bogotá - <http://www.usbbog.edu.co>

Medellín - <http://www.usbmed.edu.co>

Cali - <http://www.usbcali.edu.co>

Cartagena - <http://www.usbctg.edu.co>

Editorial Bonaventuriana - <http://www.editorialbonaventuriana.usb.edu.co/>

Revistas - <http://revistas.usb.edu.co/>

Dedicatoria

Dedico este trabajo a Dios, quien me ha dado una segunda oportunidad de vida y ha hecho un milagro en ella.

A mi familia. A la luz de mi vida, mi hija quien ha tenido toda la paciencia por el tiempo no dedicado, que me ha cuidado y acompañado, quien se alegra por todo los logros de su madre y a quien amo con mi vida

A mi madre, la mujer más maravillosa, valiente, luchadora y llena de Fe en Dios que he conocido, quien ha sido mi ejemplo a seguir porque cada día me enseña lo valiosa que es la vida y la fortaleza que debemos tener para seguir adelante, siempre con amor incondicional para su familia y ha dedicado su vida a ella.

A mi padre, quien ha sufrido en silencio todo lo que me pasado, pero siempre recordándome que hay un nuevo día y será mejor. Queriendo protegerme y cuidarme de todo siempre estando ahí con todo su amor

A mi hermana, quien ha querido cuidarme siempre y ahora más que nunca, demostrando cada día su amor desbordado, así seamos muy diferentes y no estemos de acuerdo en muchas ocasiones siempre ha estado ahí

A mi amiga Sandra Lucia, mi única amiga que me has acompañado en las buenas y en las no tantas, a quien desde conocí “siempre” ha estado conmigo dándome su apoyo, consuelo y ánimo de que podemos lograr cualquier cosa

Agradecimientos

A la Universidad de San Buenaventura seccional Cali por mostrarme que no solo es una Institución con grandes fortalezas académicas sino con un gran sentido y preocupación en la parte humana de sus estudiantes.

A todos y cada uno de los docentes con los cuales tuve la oportunidad de encontrarme en mi proceso de formación y especialmente a la profesora Rocío Emilia Segovia quien todo el tiempo estuvo a mi lado impulsándome a seguir adelante y apoyándome no solo académicamente sino personalmente, a Beatriz Grass que estuvo pendiente de mi proceso no solo académico sino también personal haciéndome sentir que podía lograrlo y que estaba con disposición de ayudarme en lo que necesitara

Tabla de contenido

I.	INTRODUCCIÓN	8
II.	CONTEXTO DE LA INVESTIGACIÓN.....	9
	A. Escenario del problema	9
	B. Marco teórico y estado del arte	9
	1) Perspectiva de Aprendizaje	9
	2) Perspectiva tecnológica	11
III.	IDENTIFICACIÓN DEL PROBLEMA	24
	A. Planteamiento del problema	24
	B. Objetivos	25
	1) General.....	25
	2) Específicos.....	25
IV.	POSTURA DE SOLUCIÓN	26
	A. Modelo de investigación	26
	1) Estrategia de investigación	27
	2) Fase Preparatoria	27
	3) Fase Descriptiva	27
	4) Fase de construcción teórica global.....	27
	5) Fase de extensión y publicación	28
V.	DESCRIPCIÓN DE LA SOLUCIÓN.....	29
	A. Diagrama de motivación	29
	B. Diagrama de implementación y migración	31
	C. Propuesta de solución.....	32
	1) Proceso Actual	32

2) Proceso con KDD	32
VI. VALIDACIÓN DE LA PROPUESTA	34
A. Plan de pruebas.....	34
B. Escenarios de pruebas	36
VII. ESTIMACIÓN Y VALORACIÓN DE IMPACTOS	38
A. Descripción del impacto esperado.....	38
1) Identificación de los elementos/circunstancias.....	38
2) Definición de aspectos relevantes al impacto esperado.....	38
3) Identificación de los elementos/circunstancias.....	39
4) Definición de aspectos relevantes al impacto esperado.....	39
VIII. CONCLUSIONES	40

LISTA DE TABLAS

Tabla I Escenarios de pruebas para el proyecto	37
Tabla II Identificación de los elementos/circunstancias	39
Tabla III Definición de aspectos relevantes al impacto esperado	39

LISTA DE FIGURAS

Fig. 1 Fases modelo CRISP-DM.....	12
Fig. 2 Actividades Comprensión del negocio	13
Fig. 3 Actividades Comprensión de los datos	14
Fig. 4 Actividades preparación de datos	15
Fig. 5 Actividades preparación de datos	16
Fig. 6 Actividades de evaluación	17
Fig. 7 Actividades de implantación.....	17
Fig. 8 Esquema del Descubrimiento de Conocimiento en Base de Datos	18
Fig. 9 Diagrama de descripción de la solución	29
Fig. 10 Diagrama de motivación.....	30
Fig. 11 Diagrama de implementación y migración	31
Fig. 12 Proceso Actual	32
Fig. 13 Proceso con KDD	33

I. INTRODUCCIÓN

Los procesos de enseñanza aprendizaje hoy en día se encuentran apoyados por plataformas tecnológicas que permiten a los estudiantes interactuar de una manera sencilla con los diferentes objetos de aprendizaje que han sido diseñados por los docentes de los cursos, sin embargo aún existen vacíos o falta de herramientas que permitan un mejoramiento en la construcción de OA que logren una mayor efectividad en el proceso de enseñanza aprendizaje.

En este documento en los capítulos I y II se contextualiza y describe la problemática existente frente a la relación entre los objetos de aprendizaje y los estilos de aprendizaje de los estudiantes y la posible solución a través de la generación de patrones de consumo de los objetos de aprendizaje para la generación de rutas de aprendizaje.

En los capítulos III y IV, se presenta la postura de solución y la descripción de la misma a la problemática evidenciada en los capítulos I y II, describiendo el diseño de la herramienta para la generación de rutas de aprendizaje a través de patrones de consumo de objetos de aprendizaje.

En los capítulos V y VI se presenta el plan para realizar la validación de la postura de solución presentada en el capítulo IV y asegurar el resultado esperado que es la obtención de las rutas de aprendizaje.

II. CONTEXTO DE LA INVESTIGACIÓN

A. *Escenario del problema*

Las nuevas modalidades de aprendizaje, tanto las totalmente virtuales como las mixtas, han traído consigo no solo una revolución en cuanto a los recursos materiales, sino que también han supuesto un cambio en la perspectiva sobre la educación (1)

En el sector de la educación se encuentran múltiples plataformas virtuales en las que se pueden implementar cursos con diversos objetos de aprendizaje y en los que toda la información tanto de los objetos como de la interacción de los estudiantes con los mismos, quedan registrados en un repositorio de información. Este repositorio de información contiene datos que si son analizados de la forma correcta sin duda favorecerían la labor del docente y mejorarían la experiencia de los estudiantes en su proceso de enseñanza- aprendizaje

Los volúmenes de información que se tienen en el repositorio de las plataformas virtuales son altísimos y su contenido valioso, para lograr tener un análisis de los datos, se deben contemplar el uso de herramientas y algoritmos que permitan hacer minería de datos y el poder generar la información necesaria para lograr obtener rutas de aprendizaje adecuadas y asociadas los estilos de aprendizaje de los estudiantes logradas a partir de identificar el consumo de objetos de aprendizaje por parte de los estudiantes.

B. *Marco teórico y estado del arte*

1) *Perspectiva de Aprendizaje*

a) *Objetos de aprendizajes*

Un objeto de aprendizaje se define como “la unidad mínima de aprendizaje, en formato digital, que puede ser reusada y secuenciada”. Se conciben, por tanto, estos pequeños componentes (OA)

como elementos integrados e integradores del proceso de enseñanza-aprendizaje, ofreciendo a los estudiantes la posibilidad de mejorar su rendimiento y nivel de satisfacción. No obstante, el OA debe cumplir una serie de características para que realmente pueda ser considerado como tal. (2)

Los objetos de aprendizajes tienen ciertas características que buscan mejorar los procesos de enseñanza- aprendizaje, una de estas es el propósito pedagógico que lo que permite es asegurar un aprendizaje satisfactorio, aunque en realidad este es el propósito esperado, pero entra a jugar un factor adicional y es los estilos de aprendizaje de cada estudiante lo que lleva a crear la necesidad de contar con herramientas que orienten a los docentes a lograr identificar ¿cuáles son los objetos de aprendizaje más adecuados para sus cursos y de acuerdo con los estilos de aprendizaje de sus estudiantes?.

El futuro no es todavía claro, pero es grande la expectativa de que los OA revolucionen los sistemas educativos. Por lo que resulta necesario prever que los OA se encuentren en condiciones de ser aprovechados adecuadamente, un requisito fundamental será tener acceso a ellos. Para su acceso, organización y reutilización, los OA se están depositando en contenedores llamados Repositorios de Objetos de Aprendizaje, los cuales permitirán que las aplicaciones, actuales y futuras, extraigan y transporten contenidos, para su modificación y reutilización. Esto beneficiará a profesores, alumnos y diseñadores, pero también será un importante soporte para el intercambio automático de contenidos entre sistemas, de un contexto de aprendizaje o de otros contextos que puedan requerirlos. (3)

Hoy en día se cuenta con diversas plataformas virtuales que permiten crear los repositorios de los OA, las que son utilizadas por estudiantes y profesores, no solo para cursos virtuales sino como apoyo a los cursos presenciales. Estas plataformas facilitan y aseguran el acceso a los contenidos, ayudas pedagógicas que permiten al estudiante a través de los OA, soportar su proceso de aprendizaje- Para los profesores la reutilización es una alternativa que facilita su actividad laboral, además que puede analizar cómo estos OA están impactando el proceso de sus estudiantes, lo que le permitirá mejorar cada día su labor en pro de sus estudiantes.

b) Estilos de aprendizaje

De acuerdo con Keefe "los estilos de aprendizaje son los rasgos cognitivos, afectivos y fisiológicos, que sirven como indicadores relativamente estables, de cómo los discentes perciben, interaccionan y responden a sus ambientes de aprendizaje" (4), para Kolb lo describe como "algunas capacidades de aprender, que se destacan por encima de otras, como resultado del aparato hereditario de las experiencias vitales propias, y de las exigencias del medio ambiente actual". (5)

Los estilos de aprendizaje son inherentes a cada individuo y están muy ligados a su interacción con el medio y la percepción que tenga de él, para lograr aprender de forma significativa. Para los docentes es fundamental conocer los estilos de aprendizajes de sus estudiantes con el fin de identificar sus preferencias frente a los diferentes OA de aprendizaje que estén asociados al curso que están realizando y que favorecen significativamente la apropiación del conocimiento.

Para la identificación de estos estilos existen diferentes modelos como : Modelo de los cuadrantes cerebrales de Herrmann, Modelo de Felder y Silverman, Modelo de Kolb, Modelo de programación neurolingüística de Bandler y Grinder , Modelo de los hemisferios cerebrales, Modelo de las inteligencias múltiples de Gardner. Estos modelos se basan en cuestionarios que se realizan a los estudiantes para identificar sus preferencias y la forma de entender el mundo, esto lleva a la identificación del estilo de aprendizaje que predomina en los actores del proceso todo con el objetivo de encontrar las mejoras formas para lograr un aprendizaje significativo e identificar otras alternativas a usar para apropiar el conocimiento

2) Perspectiva tecnológica

a) Metodologías para minería de datos

- ***Metodología Crisp DM***

Es un proceso iterativo y centrado en el negocio. Es independiente de la herramienta de Minería de datos que se utilice y está muy ligado a SPSS Clementine. (6)

CRISP-DM (Cross Industry Standard Process for Data Mining) proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos, de forma análoga a como se hace en la ingeniería del software con los modelos de ciclo de vida de desarrollo de software. El modelo CRISP-DM cubre las fases de un proyecto, sus tareas respectivas, y las relaciones entre estas tareas. En este nivel de descripción no es posible identificar todas las relaciones; las relaciones podrían existir entre cualquier tarea según los objetivos, el contexto, y el interés del usuario sobre los datos. (7)

El modelo CRISP-DM, es cíclico y consta de seis etapas: Comprender el negocio, comprender los datos, preparar los datos, modelar, evaluar e implementar. Ver Fig. 1 Fases modelo CRISP-DM

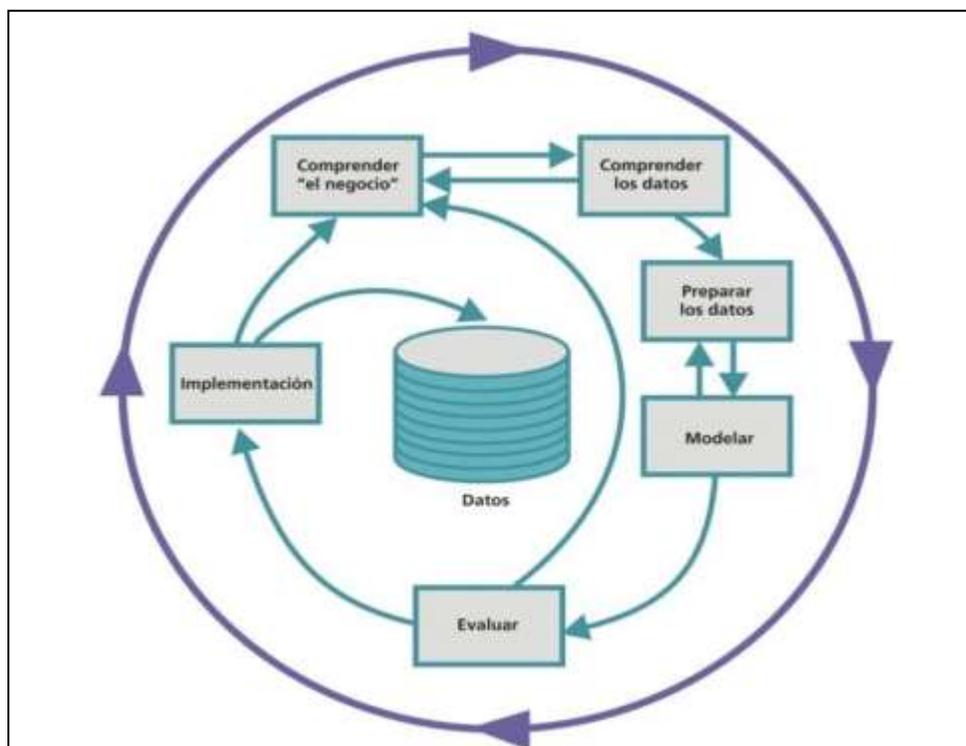
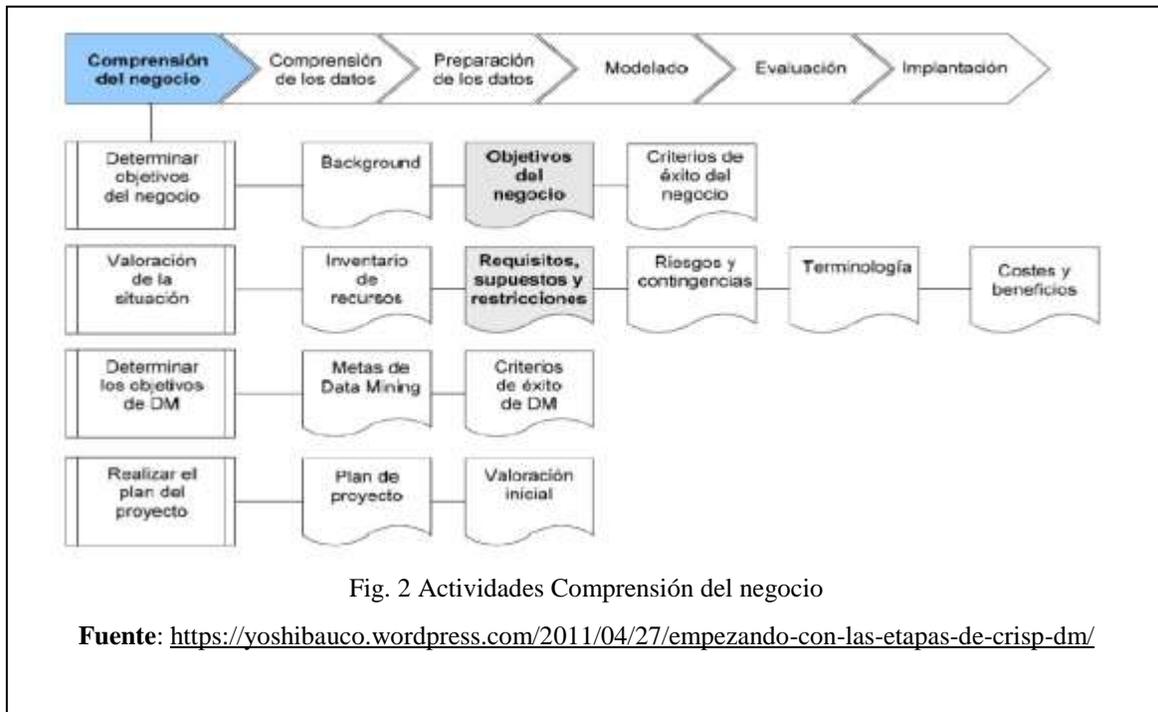


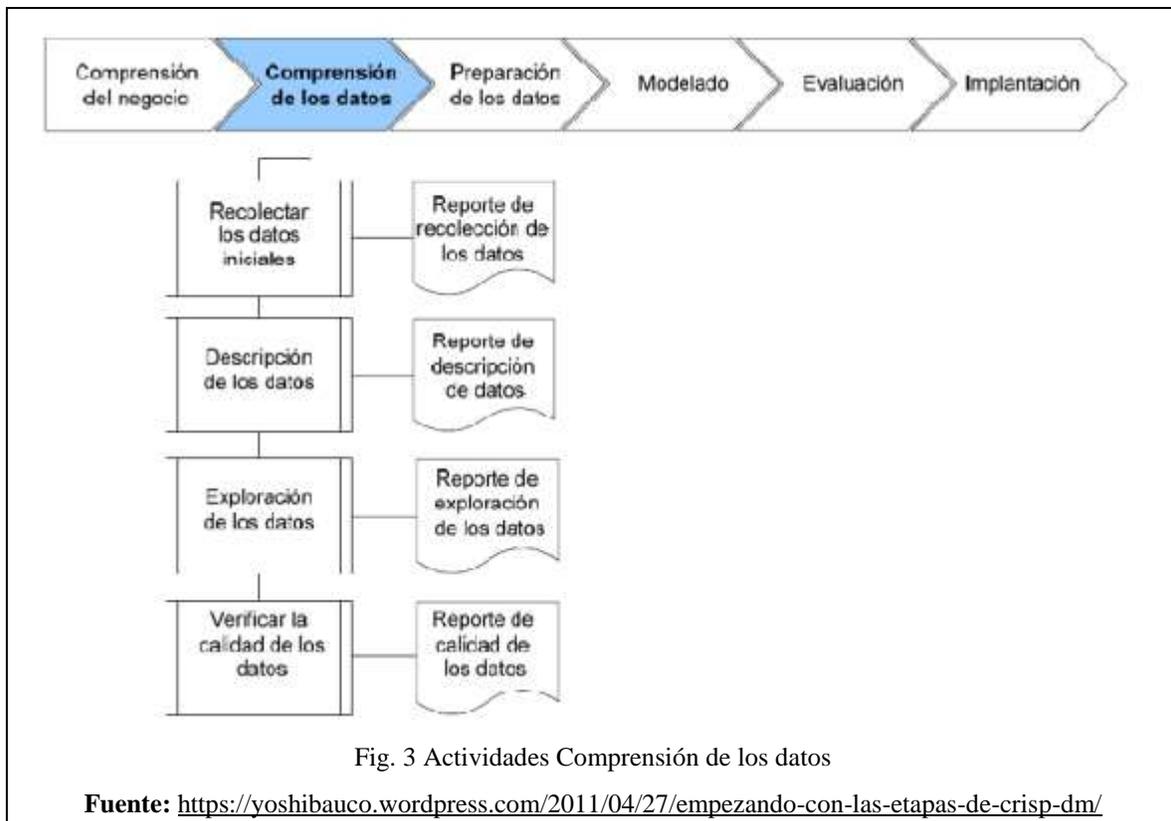
Fig. 1 Fases modelo CRISP-DM

Fuente: https://www.researchgate.net/figure/Fases-del-proceso-de-CRISP-DM-Adaptado-de-10_fig2_306959832

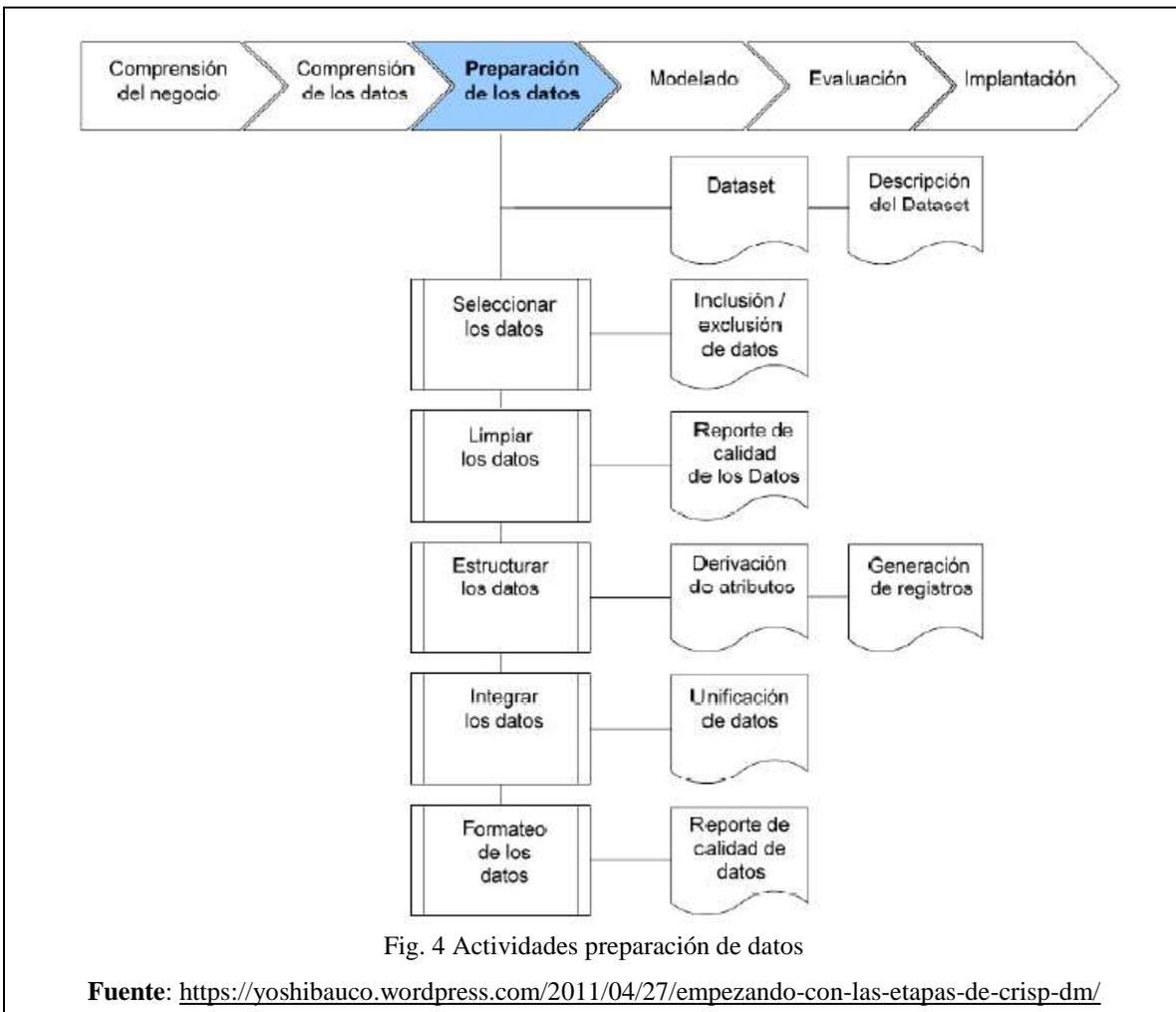
Cada fase cuenta con un conjunto de actividades, la fase I comprensión del negocio, es la fase inicial y se enfoca en la comprensión de los objetivos de proyecto. Después se convierte este conocimiento de los datos en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos. (7), ver Fig. 2. Actividades fase comprensión del negocio



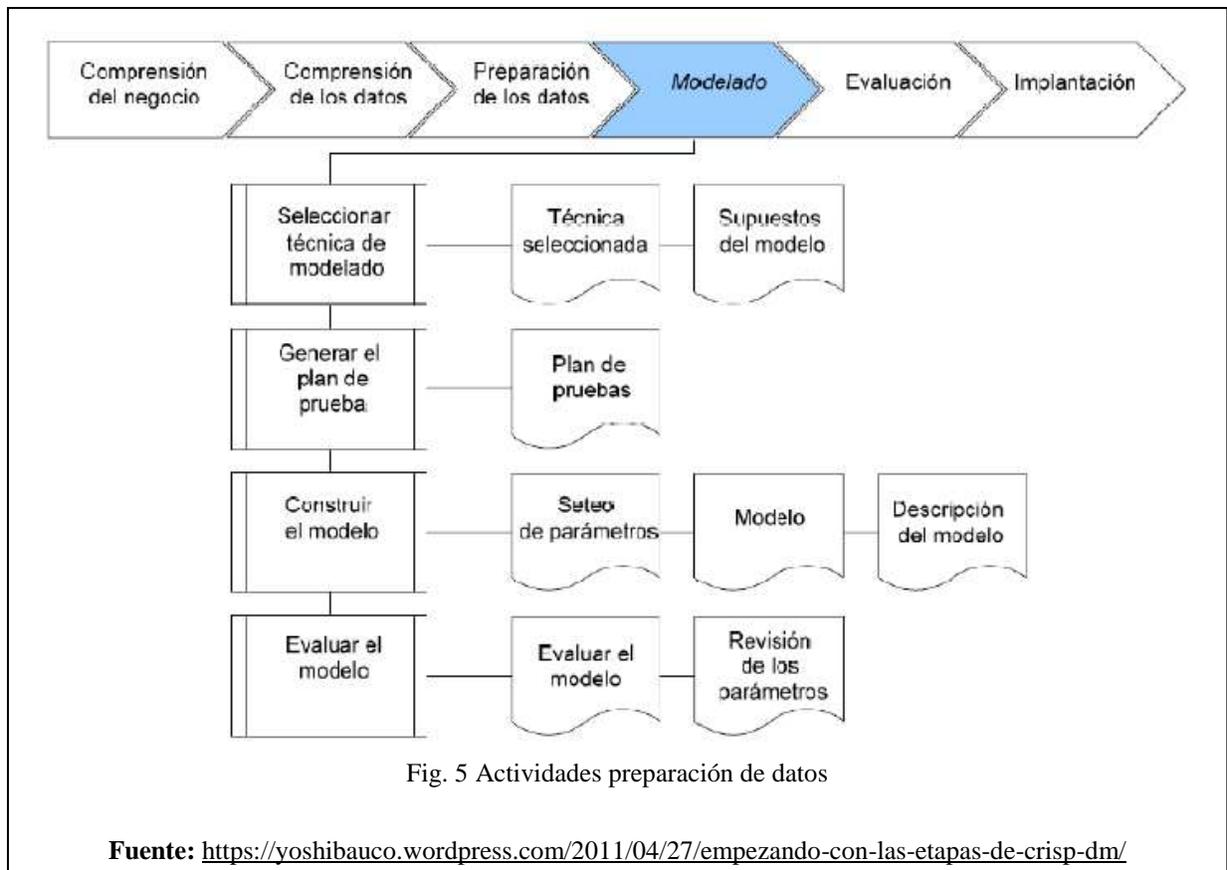
La fase de entendimiento de datos comienza con la colección de datos inicial y continúa con las actividades que permiten familiarizarse con los datos, identificar los problemas de calidad, descubrir conocimiento preliminar sobre los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta. (7), ver Fig. 3. Actividades fase comprensión de datos.



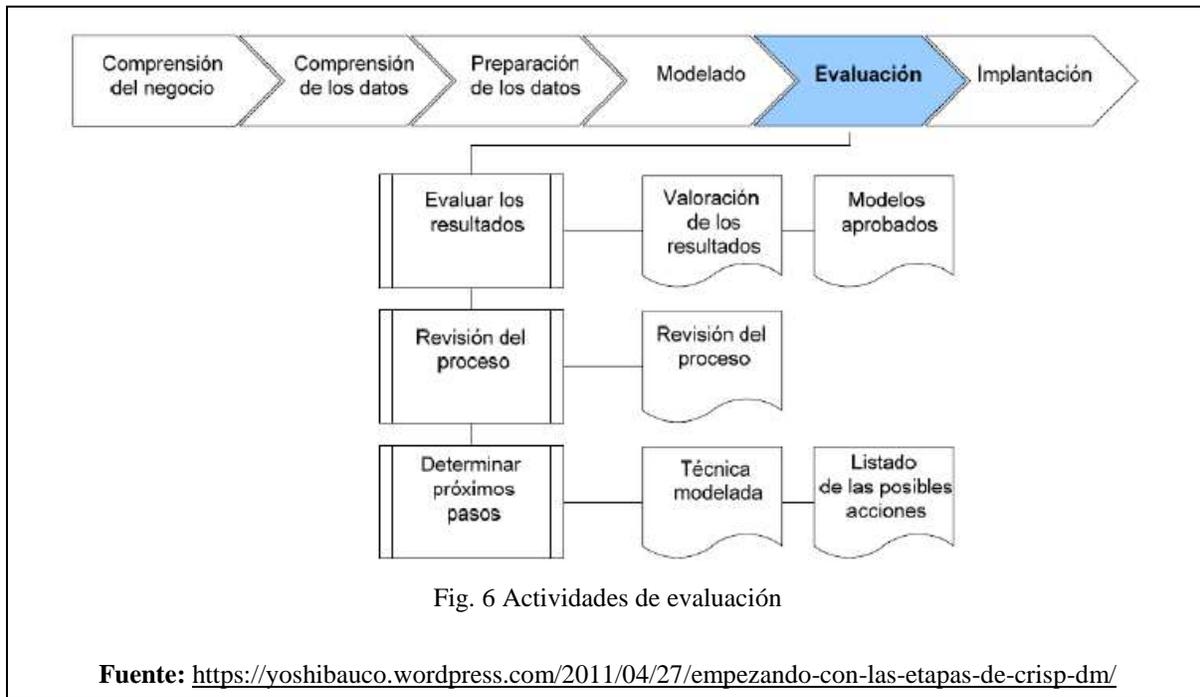
La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto final de datos (los datos que se utilizarán en las herramientas de modelado) a partir de los datos en bruto iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan. (7). Ver Fig. 4. Actividades fase preparación de datos.



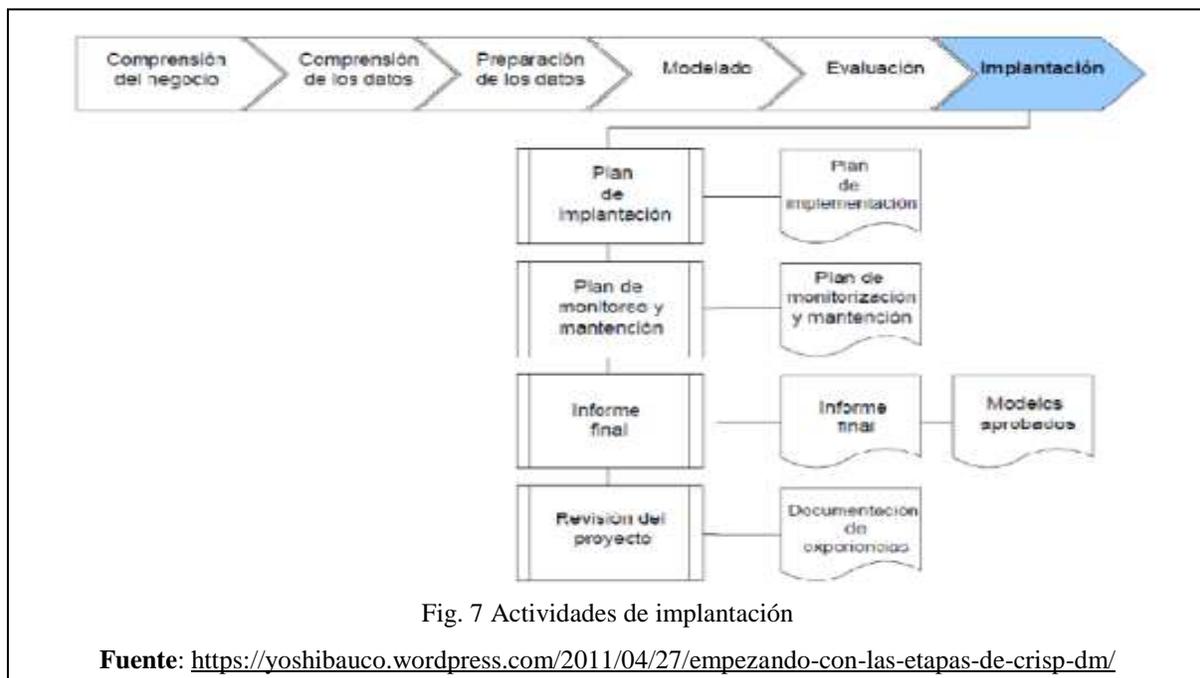
En la fase de modelado, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema (cuantas más mejor), y se calibran sus parámetros a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto se acaba volviendo a la fase de preparación de datos. Ver Fig. 5. Actividades fase modelado.



Al llegar a la fase de evaluación en el proyecto, se han construido uno o varios modelos que parecen alcanzar calidad suficiente desde la una perspectiva de análisis de datos. Antes de proceder al despliegue final del modelo, es importante evaluarlo a fondo y revisar los pasos ejecutados para crearlo, comparar el modelo obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no haya sido considerada suficientemente. Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos. (7). Ver Fig. 6 Actividades fase evaluación



La creación del modelo no es el final del proyecto, la última fase es la salida a producción. Ver Fig. 7 Actividades fase implantación



- **Metodología KDD: Proceso de Extracción de conocimiento**

La Extracción de conocimiento está principalmente relacionado con el proceso de descubrimiento conocido como Knowledge Discovery Databases (KDD), que se refiere al proceso no-trivial de descubrir conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información (8)

No es un proceso automático, es un proceso iterativo que exhaustivamente explora volúmenes muy grandes de datos para determinar relaciones. Es un proceso que extrae información de calidad que puede usarse para dibujar conclusiones basadas en relaciones o modelos dentro de los datos, ver Fig. 8 Esquema del descubrimiento de conocimiento en base de datos.

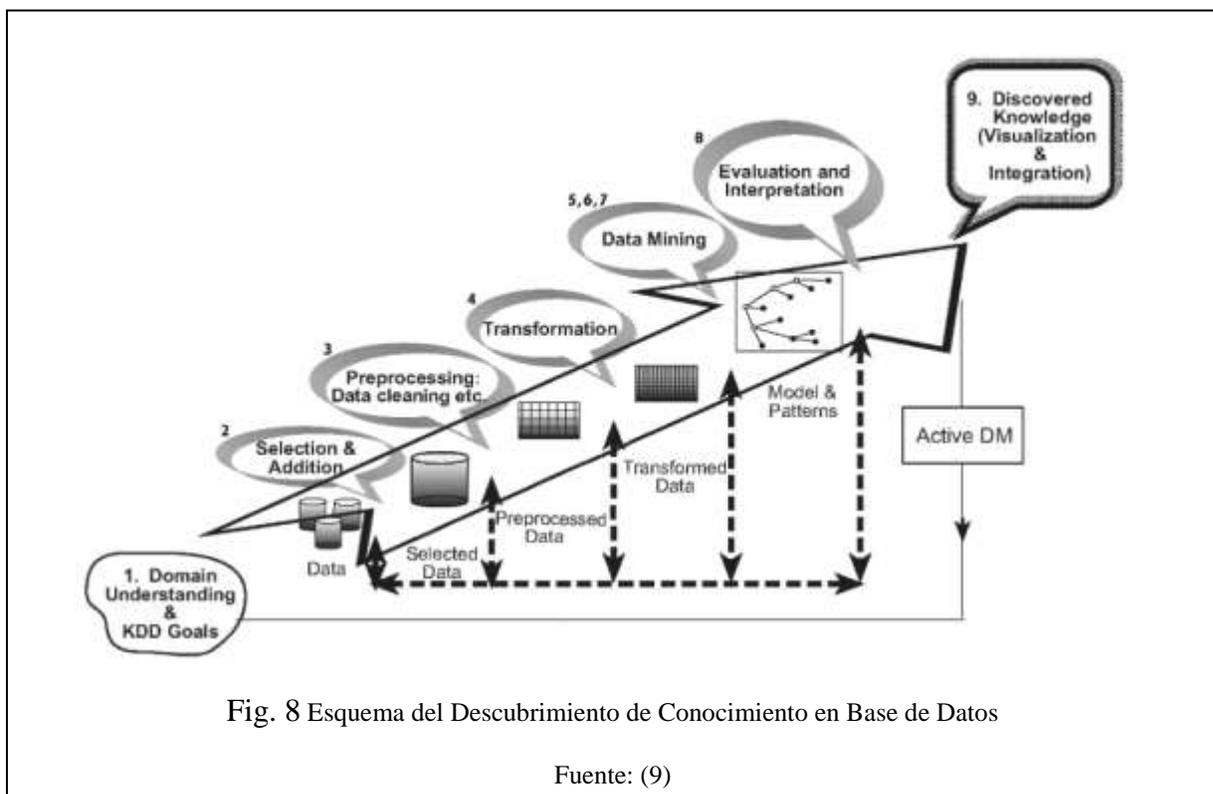


Fig. 8 Esquema del Descubrimiento de Conocimiento en Base de Datos

Fuente: (9)

Como muestra la Fig. 8, las etapas del proceso KDD se dividen en 5 fases y son:

- **Selección de datos.** En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde la o las fuentes de datos.
- **Preprocesamiento.** Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos en una forma manejable, necesaria para las fases posteriores. En esta etapa se utilizan diversas estrategias para manejar datos faltantes o en blanco, datos inconsistentes o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para su posterior transformación.
- **Transformación.** Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.
- **Data Mining.** Es la fase de modelamiento propiamente tal, en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u “ocultos” en los datos.
- **Interpretación y Evaluación.** Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos.

b) Reglas de asociación

Las reglas de asociación relacionan una determinada conclusión (por ejemplo, la compra de un producto dado) con un conjunto de condiciones (por ejemplo, la compra de otros productos). Los algoritmos de reglas de asociación buscan automáticamente las asociaciones que se podrían encontrar manualmente usando técnicas de visualización, como en el nodo Malla . (10)

Otra definición es que son similares a reglas de clasificación, pero pueden predecir cualquier atributo, cualquier combinación de atributos y no suelen utilizarse todas juntas, además que diferentes grupos muestran distintas regularidades del conjunto de datos. (11)

c) *Algoritmos de minería de datos*

Un *algoritmo* en minería de datos (o aprendizaje automático) es un conjunto de heurísticas y cálculos que permiten crear un modelo a partir de datos. Para crear un modelo, el algoritmo analiza primero los datos proporcionados, en busca de tipos específicos de patrones o tendencias. El algoritmo usa los resultados de este análisis en un gran número de iteraciones para determinar los parámetros óptimos para crear el modelo de minería de datos. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas. (12)

El modelo de minería de datos que crea un algoritmo a partir de los datos puede tomar diversas formas, incluyendo:

- Un conjunto de clústeres que describe cómo se relacionan los casos de un conjunto de datos.
- Un árbol de decisión que predice un resultado y que describe cómo afectan a este los distintos criterios.
- Un modelo matemático que predice las ventas.
- Un conjunto de reglas que describen cómo se agrupan los productos en una transacción, y las probabilidades de que dichos productos se adquieran juntos. (12)

d) *Algoritmos de búsqueda de patrones*

- **Clustering** : El Clustering (o algoritmo de agrupamiento) consiste en agrupar una serie de vectores según un criterio en grupos o clusters. Generalmente el criterio suele ser la similitud por lo que diremos que agrupa los vectores similares en grupos. Está considerado como un aprendizaje no supervisado dentro de la minería de datos. (13).

Clustering a su vez cuenta con una serie de algoritmos basados en esta filosofía estos son:

- **K-means:** Este algoritmo particiona los N objetos en K particiones (K siendo un valor arbitrario) en donde un objeto irá al clúster con la media más cercana. El algoritmo asigna K centros aleatoriamente, luego asigna los
- **Objetos al centro más cercano.** El centro se recalcula como la media de los puntos que tiene asignado, una vez actualizado se vuelven a reasignar los objetos al más cercano y así hasta tener convergencia (13).
- **K-nearest neighbors:** En este algoritmo se decide la membresía de un objeto teniendo en cuenta sus vecinos. Se decide a que clúster pertenece mirando a que clúster pertenece la mayoría de sus vecinos K más cercanos a el. Se trata de unos de los algoritmos más simples de aprendizaje automático (13)
- **Apriori:** fue uno de los primeros algoritmos desarrollados para la búsqueda de reglas de asociación y sigue siendo uno de los más empleados, tiene dos etapas:
 - Identificar todos los *itemsets* que ocurren con una frecuencia por encima de un determinado límite (*itemsets* frecuentes).
 - Convertir esos *itemsets* frecuentes en reglas de asociación. (14)

En el algoritmo Apriori se tienen un par de conceptos importantes:

- **Soporte:** El soporte del ítem o itemset X es el número de transacciones que contienen X dividido entre el total de transacciones.
- **Confianza:** La confianza de una regla “Si X entonces Y” se define acorde a la ecuación

$$\text{confianza}(X \Rightarrow Y) = \frac{\text{soporte}(\text{unión}(X, Y))}{\text{soporte}(X)} \quad (1)$$

Donde *unión* (XY) es el itemset que contienen todos los ítems de X y de Y . La confianza se interpreta como la probabilidad $P(Y/X)$, es decir, la probabilidad de que una transacción que contiene los ítems de X , también contenga los ítems de Y

En el algoritmo Apriori, encontrar *itemsets* frecuentes (*itemsets* con una frecuencia mayor o igual a un determinado soporte mínimo) no es un proceso trivial debido la explosión combinatoria de posibilidades, sin embargo, una vez identificados, es relativamente directo generar reglas de asociación que presenten una confianza mínima. El algoritmo *Apriori* hace una búsqueda exhaustiva por niveles de complejidad (de menor a mayor tamaño de *itemsets*). Para reducir el espacio de búsqueda aplica la norma de “si un *itemset* no es frecuente, ninguno de sus *supersets* (*itemsets* de mayor tamaño que contengan al primero) puede ser frecuente”. Visto de otra forma, si un conjunto es infrecuente, entonces, todos los conjuntos donde este último se encuentre, también son infrecuentes. (14)

- **Fp-growth:** Este algoritmo fue propuesto en el año 2000, y lo que hace es que permite extraer reglas de asociación a partir de *itemsets* frecuentes pero, a diferencia del algoritmo Apriori estos se identifican sin necesidad de generar candidatos para cada tamaño. El algoritmo emplea una estructura de árbol (Frequent Pattern Tree) donde almacena toda la información de las transacciones. Esta estructura permite comprimir la información de una base de datos de transacciones hasta 200 veces, haciendo posible que pueda ser cargada en memoria RAM. Una vez que la base de datos ha sido comprimida en una estructura FP-Tree, se divide en varias bases de datos condicionales, cada una asociada con un patrón frecuente. Finalmente, cada partición se analiza de forma separada y se concatenan los resultados obtenidos. En la mayoría de casos, FP-Growth es más rápido que Apriori. (14)

e) *Herramientas Tecnológicas de DM*

- **KNIME** (o Konstanz Information Miner) es una plataforma analítica que permite el desarrollo de modelos en un entorno visual. Está construido sobre la plataforma Eclipse. y es extremadamente flexible y potente. Bajo una filosofía parecida a la de PDI, KNIME permite

manipular datos , analizarlos, realizar flujos de datos en base a un modelo de nodos que se conectan entre sí. (15)

Es una plataforma modular de exploración de datos, permite al usuario crear flujos de datos, además permite ejecutar de forma selectiva algunos de los pasos creados, así como ejecutar todo el flujo desarrollado. Después de la ejecución, los resultados se pueden revisar mediante vistas interactivas tanto de los datos como de los modelos (15)

El proceso de análisis de datos, utilizado por KNime, consiste de un pipeline de nodos, conectados a través de puentes que transportan datos o modelos. Cada nodo procesa la llegada de datos y/o modelo(s) y produce resultados como salida. (16)

De acuerdo con (17) la herramienta Knime puede ser aplicada en:

- Tareas de preprocesamiento de datos o ETL (Extract, transform and load) soportando diversas fuentes de datos como xls (Excel), xml, Json o bases de datos.
- Paralelización de algoritmos (Integrable con Hadoop y Spark)
- Análisis estadístico (básico y avanzado)
- Análisis de comunidades, grafos y redes sociales, incluyendo reporting (reportería)
- Procesamiento del lenguaje natural o Text mining (Minería de texto) y web scraping
- Procesamiento de imágenes
- Análisis químico y molecular (también empleado en bioinformática)

III. IDENTIFICACIÓN DEL PROBLEMA

A. *Planteamiento del problema*

La tecnología forma parte fundamental de los procesos de enseñanza aprendizaje de las instituciones de educación, hoy en día se cuenta con una gran variedad de plataformas virtuales que les permiten a los estudiantes tener un acceso a los diferentes contenidos y objetos de aprendizaje que suben los profesores en cada uno de los cursos. A pesar de esto no se cuenta con las herramientas que permitan a los profesores hacer una identificación de las preferencias de los estudiantes frente a los diferentes objetos de aprendizaje y mucho menos con cuales se ha logrado un cumplimiento de los objetivos de cada uno de los cursos.

La construcción de objetos de aprendizaje por parte de los profesores para los cursos son definidos por ellos mismos de acuerdo a sus propios skills y a su propia forma de aprendizaje e incluso orientados a sus gustos personales o facilidad en la construcción y no son el resultado de un análisis de información que permita identificar si realmente han cumplido el objetivo de lograr facilitar el proceso de enseñanza aprendizaje y que este sea significativo para los estudiantes.

Frente a la utilización de OA en plataformas virtuales y el cumplimiento real del objetivo de los mismo se ha identificado que:

- En un bajo porcentaje se hace la aplicación de algún modelo de estilo de aprendizaje para que este sea identificado en los estudiantes antes de iniciar un curso.
- La construcción de los OA a ser abordados por los estudiantes en los cursos no son diseñados basados en una análisis del estilo de aprendizaje de quienes los están abordado.
- No se logra identificar la prioridad, ni la frecuencia de uso de estos OA después de finalizado el curso.

- No se logra identificar la efectividad y que tan significativos han sido los OA para el proceso de apropiación de conocimientos del estudiante

Por lo descrito anteriormente se hace necesario contar con una herramienta que permita identificar patrones de consumo de los OA de cursos que se encuentran en plataformas virtuales , para hacer un análisis acerca de la forma de aprender de quienes los usan y que tan efectivo y significado han sido en su proceso con el fin de generar rutas de aprendizaje en los cursos que se encuentran desarrollando, logrando apoyar a los docentes en su labor.

B. Objetivos

1) General

Diseñar una herramienta de identificación de rutas de aprendizaje basado en patrones de consumo de OA en plataforma MOODLE

2) Específicos

- Identificar algoritmos que permitan obtener patrones de consumo de OA.
- Identificar herramientas para ejecución de los algoritmos de patrones.
- Diseñar una ETL a partir del repositorio de datos de moodle hacia una tabla minable.
- Diseñar una herramienta de visualización de rutas de aprendizaje a través de los patrones de consumo.
- Plantear los escenarios de pruebas que permitan validar la solución.

IV. POSTURA DE SOLUCIÓN

El diseño de una herramienta de identificación de rutas de aprendizaje basado en patrones de consumo de Objetos de aprendizaje en la plataforma MOODLE, es una alternativa innovadora que apoyará el proceso de enseñanza-aprendizaje porque permitirá que los docentes tengan la información necesaria acerca de la interacción que tienen los estudiantes con los OA de sus cursos.

Los docentes podrán analizar la información y apoyarse en ella para la construcción y mejora de los OA que tienen en las plataforma, además de poder analizar que rutas de aprendizajes son las más adecuadas de acuerdo con el estilo de aprendizaje de sus estudiantes, esto favorecería el aprendizaje de los estudiantes.

A. *Modelo de investigación*

Para el desarrollo de este proyecto se revisaron las metodologías más utilizadas en procesos de minería de datos y que permita la construcción de patrones, este proyecto se apoyará en fases de la metodología de minería de datos KDD. Además se hará uso del modelo de datos de la plataforma virtual MOODLE

Con este marco de trabajo y la fuente de información se logrará obtener una propuesta de solución para dar cumplimiento a los objetivos

La propuesta está compuesta de una serie de fases que al ser desarrolladas permitirá obtener un diseño de una herramienta que en trabajos futuros podrá ser implementada para lograr un mejoramiento en la labor docente frente a la construcción de OA e identificación de las mejora rutas de aprendizaje para sus estudiantes y en el caso de estudiantes el uso de OA cumplirá de forma más efectiva su objetivo.

1) *Estrategia de investigación*

La metodología apropiada para este proyecto es KDD para minería de datos, para ésta investigación aplicada, se empleará el “Modelo de investigación Documental”, cuyas fases son:

2) *Fase Preparatoria*

- Se identificarán los conceptos desde las perspectivas del aprendizaje y la tecnología.
- Metodologías
- Algoritmos de patrones de consumo
- Herramientas para la ejecución de algoritmos de patrones de consumo

3) *Fase Descriptiva*

- Revisión del repositorio de datos de Moodle
- Revisión de proyectos similares con respecto a la construcción de rutas de aprendizaje e identificación de patrones de consumo asociados a los procesos educativos

4) **Fase de construcción teórica global**

- Conjunto de resultados del estudio adelantado
- Identificación de vacíos
- Limitaciones
- Dificultades
- Tendencias

- Logros esperados durante el desarrollo del proyecto

5) **Fase de extensión y publicación**

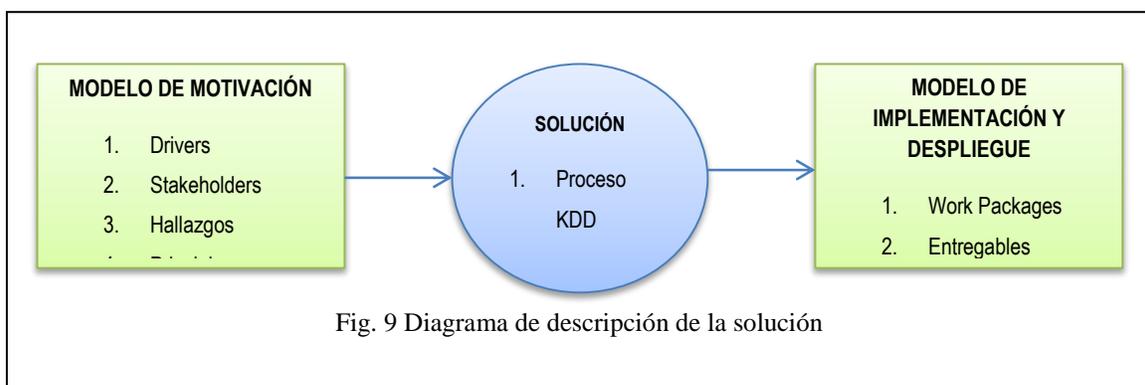
- Divulgación de los resultados obtenidos

V. DESCRIPCIÓN DE LA SOLUCIÓN

Para la descripción de la solución se emplean conceptos del marco de trabajo TOGAF para Arquitectura Empresarial y la metodología KDD. La metodología ADM de TOGAF ofrece la manera cómo implementar arquitectura de TI, apoyada en el lenguaje de diagramación Archimate.

La descripción de la solución, ver Fig. 9 diagrama de descripción de la solución, consta de las siguientes partes:

- Modelo de motivación
- Solución propuesta



Los modelos ADM a ser usados, corresponden a los de Motivación e implementación y despliegue

A. *Diagrama de motivación*

El diagrama de motivación estructura jerárquicamente los interesados (Roles y Stakeholders), las necesidades evidenciadas (Drivers), los hallazgos (Assessment), los objetivos (Goal) a ser logrados y la manera como serán atendidos (Requirements). Ver Fig. 10 Diagrama de Motivación.

De igual manera se incluyen los servicios que se evidencian para ser usados en la solución, así como los productos de software que soportarían dichos servicios.

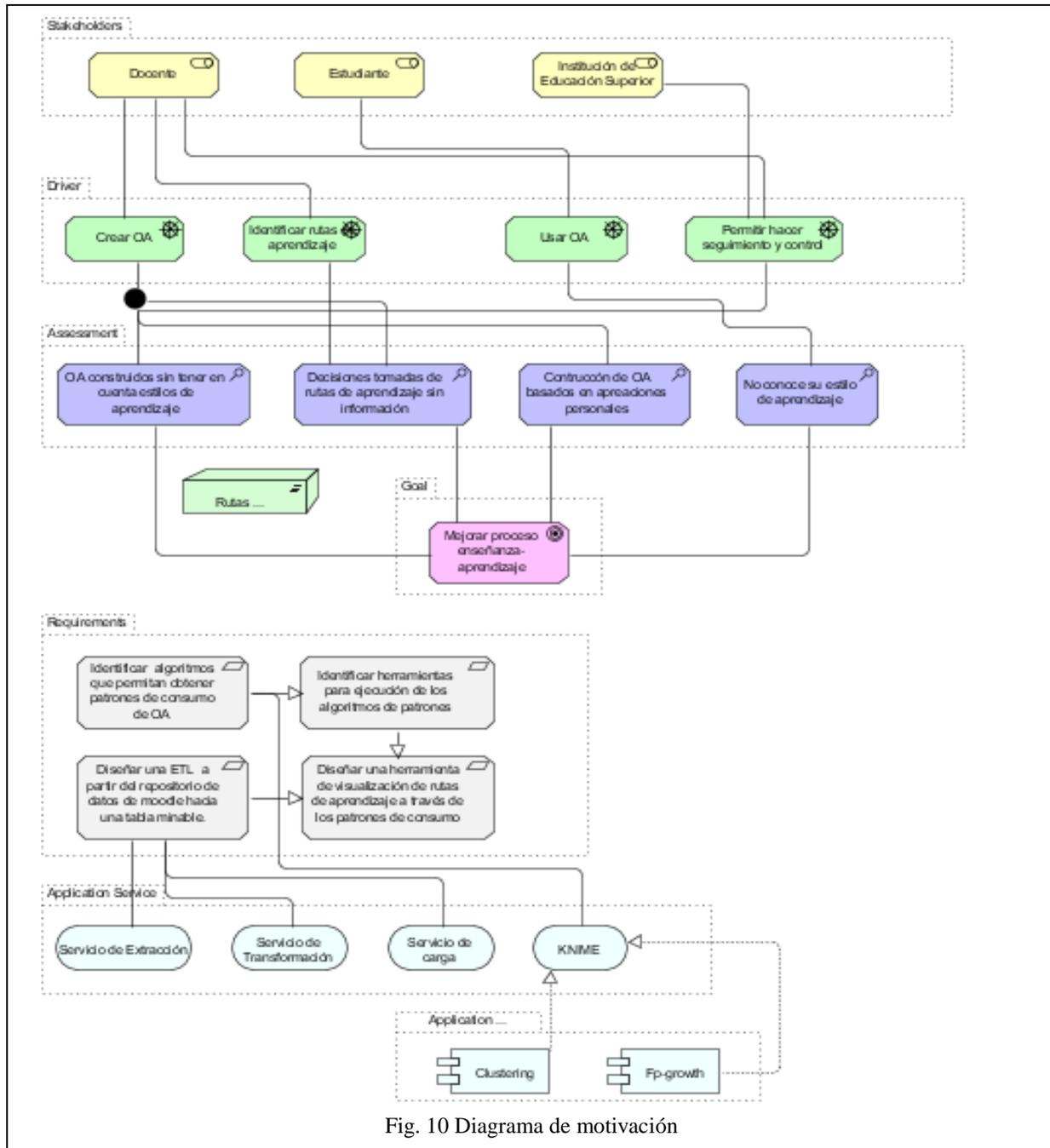


Fig. 10 Diagrama de motivación

B. Diagrama de implementación y migración

El diagrama de implementación y migración permite mostrar la estructura de la solución y la manera como será puesta en servicio, teniendo en cuenta las brechas (gaps) que deben ser analizadas y con qué paquete de trabajo serán cubiertas y a su vez estos paquetes de trabajo que objetivos son los que van a responder. Ver Fig. 11 Diagrama de implementación y migración

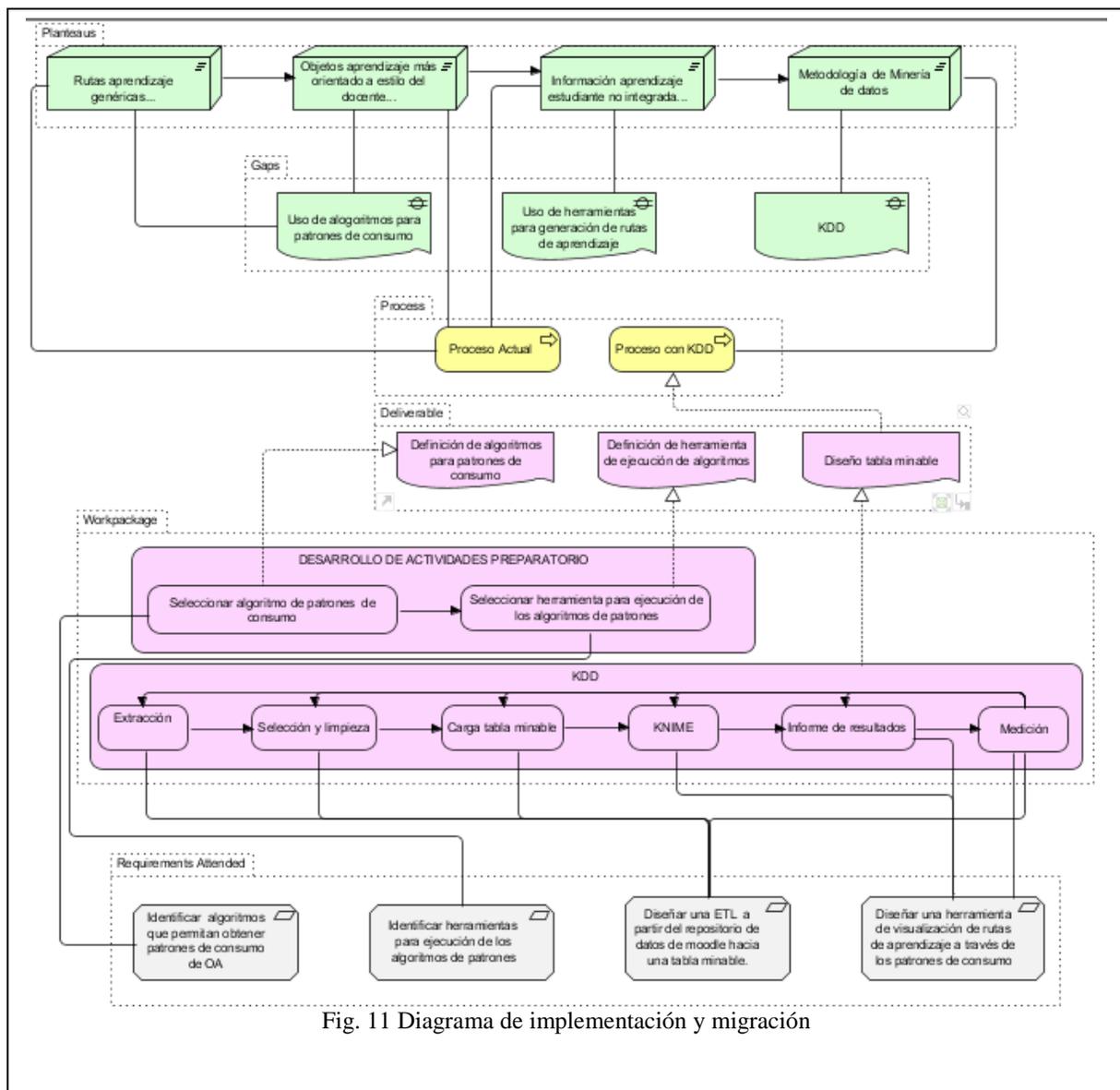


Fig. 11 Diagrama de implementación y migración

Se proyecta abordar la implementación haciendo uso de lo existente en el proceso base, específicamente el repositorio de datos de moodle que será el insumo para el proceso de minería de datos que se desarrollará con la metodología KDD, para la identificación de los patrones de consumo de OA se utilizarán los algoritmos de clustering y Fp-growth a través de la herramienta KNIME, la que se encargará de ejecutar los algoritmos que generaran la rutas de aprendizaje.

C. *Propuesta de solución*

Tal como está planteado en el diagrama de Implementación y Migración, la solución consta de una fase intermedia donde se acogen los lineamientos del modelo base y buenas prácticas de gestión de datos, y una fase final, a la cual se le agregan procesos de gestión de metadatos. Se presentan a continuación los diagramas correspondientes

1) *Proceso Actual*

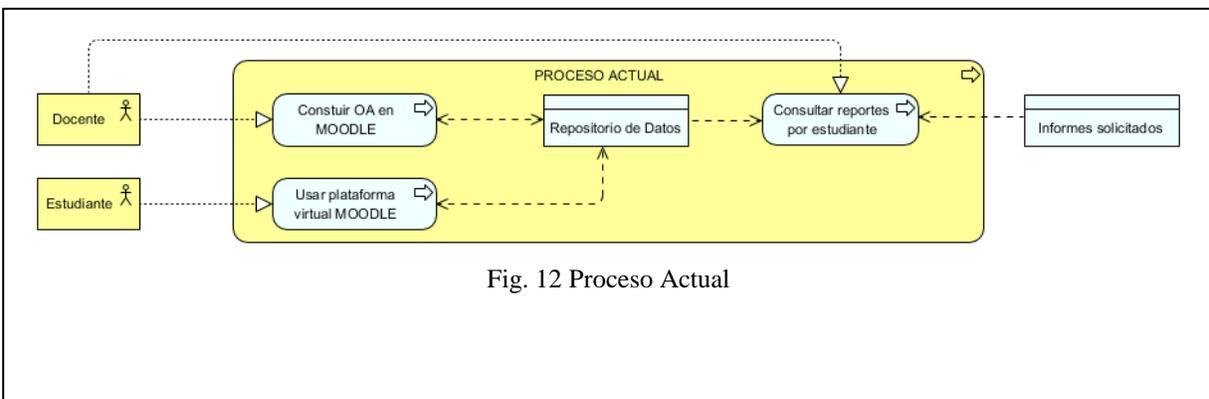


Fig. 12 Proceso Actual

2) *Proceso con KDD*

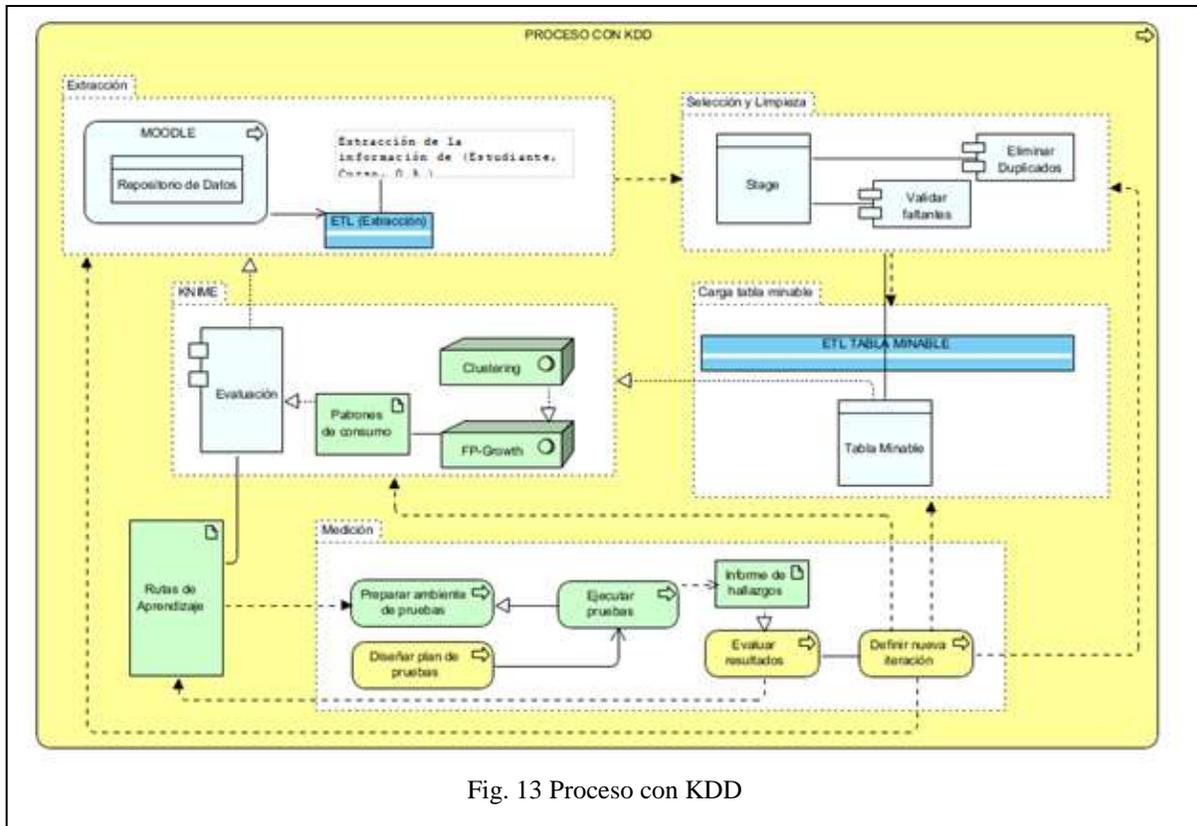


Fig. 13 Proceso con KDD

VI. VALIDACIÓN DE LA PROPUESTA

A. *Plan de pruebas*

El objetivo de ésta sección es definir el alcance del plan de pruebas a ser realizadas para validar el proceso de minería de datos utilizando la metodología KDD.

Aspectos generales para la ejecución de los escenarios de pruebas:

El enfoque de las pruebas es por las actividades que se deben ejecutar en el desarrollo de la metodología KDD, dado que se requiere validar si cada una de las actividades está cumpliendo con lo requerido como insumo de la actividad siguiente y si la ejecución de los algoritmos en la herramienta KNIME, está arrojando la información esperada acerca de los patrones de consumo de objetos de aprendizaje para la consecución de las rutas de aprendizajes.

Escala de valores: para la realización de las pruebas se considerará la siguiente escala de valores

- **No cumple (NC):** si para todos los casos de prueba asociados al escenario no se evidencia su cumplimiento.
- **Cumple Parcialmente (CP):** Si cumple en algunos casos de prueba asociados al escenario y en otros no.
- **Si cumple (SC):** Si para todos los casos de prueba asociados al escenario considerados, se evidencia su cumplimiento.

Ambiente de pruebas: El ambiente de pruebas debe integrarse desde la primera fase de la metodología KDD con el fin de lograr el aseguramiento de la data extraída para las actividades posteriores, lograr identificar si se requiere hacer la actividad de selección y limpieza de datos, además asegurar que la carga de data para la tabla minable esté completa de tal manera que los datos utilizados en el proceso de minería sean confiables para obtener los resultados esperados.

Iteraciones requeridas en la ejecución de pruebas: La cantidad de veces requeridas para asegurar el resultado depende de los resultados en la actividad de medición esto ocasionado porque allí se debe evaluar los resultados del proceso de minería, se realiza la retroalimentación necesaria a cada actividad y se debe volver a iterar en caso de encontrar hallazgos. De tal manera que el número de iteraciones depende del resultado de la medición, este resultado debe ser recreado en un informe de hallazgos que permita identificar exactamente la actividad y las necesidades de información requeridas para dar cumplimiento a los objetivo de generar las rutas de aprendizaje de acuerdo con los patrones de consumo de los OA por parte de los estudiantes que acceden a los cursos en las plataformas virtuales.

La ejecución de escenarios tiene como objetivo principal la detección de fallas en la atención de los hallazgos, en tal sentido, las actividades macro orientadas a mejorar evidencias detectadas están relacionadas con:

- Preparar el ambiente en el cual la prueba va a ser realizada
- Diseñar los casos de prueba para cada escenario
- Ejecutar los casos de prueba correspondiente al escenario
- Documentar el nivel de severidad del no cumplimiento de la prueba.
- Determinar el componente que genero la situación anómala
- Proceder a generar la medida para solucionar la falla
- Aplicar la solución generada
- Re-agendar la realización de nuevas pruebas

Los hallazgos presentan diferentes niveles de criticidad por lo cual deben ser atendidos en el siguiente orden:

- **Modelado tabla minable inadecuado:** Este hallazgo es el principal porque no permite contar con los campos necesarios y adecuados que lleven a lograr la generación de los patrones de consumo y lograr la identificación de las rutas de aprendizaje.

- **Limpieza y selección sin análisis:** Ocasiona inconsistencias en al data como duplicidad o faltantes lo que lleva errores en la generación de los patrones de consumo de los OA
- **Extracción:** Ocasiona tener los registros incompletos para la carga en la tabla minable, situación que lleva a no tener la data completa para la generación de los patrones de consumo de los OA
- **Carga de datos tabla minable incompleto:** El que se presenten hallazgos en la carga impide tener los registros necesarios para la identificación de las rutas de aprendizaje.

Los indicadores que permitan identificar en estado de no cumplimiento se evaluarán con la siguiente medición:

- **No cumple:** si se evidencia su incumplimiento
- **Cumple Parcialmente:** Si cumple en algunos casos y en otros no.
- **Si cumple:** Si para todo No se evidencien a partir del uso del modelo s los casos considerados se evidencia su cumplimiento

B. Escenarios de pruebas

La obtención de la medición se lleva a cabo mediante el producto de la frecuencia de los errores observados por el impacto de dicho hallazgo.

El impacto se valorará de la siguiente manera:

- Bajo: 20
- Medio: 50
- Alto: 80

Tabla I Escenarios de pruebas para el proyecto

Atributo	Componente	Capacidad	Impacto	Medición
Exactitud Recuperabilidad	Extracción	Identificar OA con : Id_curso, tipo OA, fecha uso, hora de uso, id_estudiante, nota	Alto	IND = FE * INP
Consistencia Precisión	Limpieza de datos	Datos limpios y sin faltantes ni duplicados	Alto	IND = FE * INP
Compleitud	Modelo	Tabla minable con toda la información de los estudiantes donde se identifica el conjunto de OA consultados por los estudiantes por transacción	Alto	IND = FE * INP
Precisión	KNIME	Identificar patrones de consumo a los que se pueda relacionar y clasificar con los estilos de aprendizaje	Alto	IND = FE * INP

VII. ESTIMACIÓN Y VALORACIÓN DE IMPACTOS

El análisis de impactos busca determinar los logros que el proyecto permitirá alcanzar, relacionados con la atención de los hallazgos levantados, así como los aspectos adversos que pudiesen afectar la obtención de dichos logros.

A. *Descripción del impacto esperado*

El impacto general del proyecto es “Contar con el diseño una herramienta de identificación de rutas de aprendizaje basado en patrones de consumo de OA en plataforma MOODLE” el cual permita a al equipo docentes de una institución que usa Moodle como plataforma virtual mejorar sus objetos de aprendizaje y asociarlo al estilo de aprendizaje de sus estudiantes de una forma más asertiva como se han evidenciado en el diagrama de Motivación.

1) *Identificación de los elementos/circunstancias*

La implementación del proyecto se logrará con un trabajo futuro que desarrolle la solución diseñada en esta propuesta. Esta herramienta es una alternativa innovadora para el proceso educativo de cualquier institución porque sin duda mejorará la interacción del estudiante con los OA y las plataformas virtuales y logrará tener un aprendizaje más efectivo

2) *Definición de aspectos relevantes al impacto esperado*

En la adopción del proceso propuesto, se presentan ciertas situaciones que pueden atentar contra su implementación, y debería entonces diseñarse un plan que permita mitigar las causas relacionadas.

Los principales riesgos evidenciados tienen que ver con el rechazo al proceso propuesto por parte de los docentes frente a la inminente necesidad de ajustar los OA de acuerdo a los estilos de aprendizaje de los estudiantes porque ya se contaría con rutas de aprendizaje definidas.

3) *Identificación de los elementos/circunstancias*

Tabla II Identificación de los elementos/circunstancias

Id Impacto	Tipo de Impacto	Descripción del impacto
IMP01	Tecnológico	Se potencializa el uso de la plataforma virtual que utiliza la institución de educación superior para sus cursos y se abre espacio a la implementación de nuevas herramientas para la minería de datos que ofrecen un mayor conocimiento de las necesidad que tiene cada estudiante
IMP02	Eficiencia en Procesos	El uso herramientas tecnológicas que ofrezca la información necesaria para analizar y mejorar los procesos de enseñanza aprendizaje logrando hacer más efectivo la apropiación del conocimiento para los estudiantes, sin lugar a dudas harán más eficientes los procesos.
IMP03	Imagen Corporativa	El que los estudiantes interactúen con OA que se adapten a su estilo de aprendizaje y que sientan que su proceso es efectivo, sin lugar a dudas lleva a que la imagen de la institución educativa en la que están desarrollando su proceso la consideren el mejor lugar para estudiar y sea muy positiva .

4) *Definición de aspectos relevantes al impacto esperado*

Tabla III Definición de aspectos relevantes al impacto esperado

Id	Afectación	Consideraciones de afectación	Estrategia de Mitigación
IMP02	Rechazo del proceso propuesto por parte de los docentes	El tener docentes que han llevado cursos durante muchos años y haciendo uso de los mismos materiales y objetos de aprendizaje genéricos, podrían afectar la aceptación de la iniciativa:	Gestionar el cambio con capacitaciones en construcción de objetos de aprendizaje asociado a los estilos de aprendizaje

VIII. CONCLUSIONES

- Los procesos de enseñanza aprendizaje cada día requieren ofrecer formas más innovadoras y efectivas para la apropiación del conocimiento por parte de los estudiantes, para tal fin se están apoyando en plataformas tecnológicas y objetos de aprendizaje virtuales que son genéricos y no están asociados en realidad a los estilos de aprendizaje de los estudiantes activos en cada curso.
- Existe una necesidad inminente de mejorar los procesos de enseñanza aprendizaje y de contar con herramientas que ofrezcan mayor información al docente para la construcción de objetos de aprendizaje que aseguren la efectividad de los mismos en el proceso individual de cada estudiante vinculado a los cursos
- Las rutas de aprendizaje son una alternativa para que los docentes logren identificar como a sus estudiantes se les facilitaría el aprendizaje y les permitiría construir objetos de aprendizaje que sean muy efectivos para la apropiación por parte del estudiante de los contenidos de los cursos.
- La identificación de patrones de consumo de objetos de aprendizaje son la herramienta adecuada para lograr obtener rutas de aprendizaje asertivas y lograr el objetivo que es tener OA efectivos para los estudiantes
- La minería de datos , las plataformas que permiten ejecutar algoritmos y las metodologías para el desarrollo de este proceso son la mejor alternativa para el tratamiento del volumen de data que se genera en las plataformas virtuales de las instituciones educativas y permitirán tener la consolidación de información para que se realicen los análisis de las mismas en pro del mejoramiento de los procesos educativos.
- Este proyecto es un insumo para el desarrollo e implementación de la herramienta de herramienta de identificación de rutas de aprendizaje basado en patrones de consumo de Objetos de aprendizaje en la plataforma MOODLE

REFERENCIAS

1. AGUÑA GA. Patrones en aprendizaje: Concepto, aplicación y diseño de un patrón. RED – Revista de Educación a Distancia. Número 31. 2012;; p. 1.
2. Vicerrectorado de Estudios y Convergencia Europea. Plan de acciones para la convergencia Europea. 2006..
3. López C. Los repositorios de objetos de aprendizaje como soporte a un entorno e-learning, Tesina doctora, Universidad de Salamanca; 2005.
4. Keefe JW,MJS. Learning Styles Profile Examiner’s, Reston, VA:NASS. 1986..
5. Kolb A KD. Learning Styles and Learning Spaces: Enhancing Experimental Learning in Higher Education. Academy of Management Learning & Education. 2005..
6. Big Data International Campus. Big Data International Campus. [Online].; 2018 [cited 2018 Agosto 30. Available from: HYPERLINK "https://www.campusbigdata.com/big-data-blog/item/104-procesos-de-analisis-de-datos" <https://www.campusbigdata.com/big-data-blog/item/104-procesos-de-analisis-de-datos> .
7. Villena RJ. Singular data & analytics. [Online].; 2016 [cited 2018 Agosto 30. Available from: HYPERLINK "https://data.sngular.team/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science" <https://data.sngular.team/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science> .
8. Han J, M K. Data Mining: Concepts ans Techniques USA: Morgan Kaufmann Publishers; 2001.
9. Oded M, Lior R. Data Mining and Knowledge Discovery Handbook New York: Srpinger-Verlag; 2005.
10. IBM. IBM Knowledge Center. [Online].; 2016. Available from: HYPERLINK "https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/nodes_associationrules.html" https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/nodes_associationrules.html .
11. González CA. Departamento de Informática - Universidad de Valladolid. [Online].; 2015 [cited 2018 Agosto 30. Available from: HYPERLINK "https://www.infor.uva.es/~calonso/MUITIC/MineriaDatos/04InduccionReglasAsociacion.pdf" <https://www.infor.uva.es/~calonso/MUITIC/MineriaDatos/04InduccionReglasAsociacion.pdf> .
12. Duncan O. Microsoft Docs. [Online].; 2017 [cited 2018 Agosto 30. Available from: HYPERLINK "https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=sql-server-2017" <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=sql-server-2017>

- [services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=sql-server-2017](#) .
13. Blanco EJ&SH. Universidad Politecnica de Cataluña. [Online].; 2016. Available from: HYPERLINK "https://upcommons.upc.edu/bitstream/handle/2117/82434/113257.pdf?sequence=1&isAllowed=y" <https://upcommons.upc.edu/bitstream/handle/2117/82434/113257.pdf?sequence=1&isAllowed=y> .
 14. Amat J. RPubS. [Online].; 2018 [cited 2018 Agosto 30. Available from: HYPERLINK "https://rpubs.com/Joaquin_AR/397172" https://rpubs.com/Joaquin_AR/397172 .
 15. Jortilles. Jortilles. [Online].; 2017 [cited 2018 Agosto 30. Available from: HYPERLINK "http://mineriadedatosknime.blogspot.com/" <http://mineriadedatosknime.blogspot.com/> .
 16. Nigro O. Facultad de Ciencias Exactas- Universidad Nacional del Centro de la Provincia de Buenos Aires. [Online].; 2012 [cited 2018 Agosto 30. Available from: HYPERLINK "http://www.exa.unicen.edu.ar/catedras/dmining/clases/PresentacionKNime.pdf" <http://www.exa.unicen.edu.ar/catedras/dmining/clases/PresentacionKNime.pdf> .
 17. Stradata. Stradata. [Online].; 2016 [cited 2018 Agosto 30. Available from: HYPERLINK "https://stradata.co/primera-experiencia-con-knime/" <https://stradata.co/primera-experiencia-con-knime/> .