

# Test Modeling for BI (Business Intelligence) Projects

*Katerine Villamizar Suaza, Magister en Ingeniería de Sistemas<sup>1</sup>, Dario Enrique Soto Durán, Magister Ciencias de la Computación<sup>2</sup>, Juan Camilo Giraldo Mejía, Magister en Ingeniería de Sistemas<sup>3</sup>, Jovani Alberto Jiménez Builes, Doctor en Ingeniería de Sistemas<sup>4</sup>,  
<sup>1</sup>Tecnológico de Antioquia, Colombia, kvillamizar@tdea.edu.co; <sup>2</sup>dsoto@tdea.edu.co; <sup>3</sup>jgiraldo1@tdea.edu.co. <sup>4</sup>Universidad Nacional de Colombia, jajimen1@unal.edu.co.*

**Abstract-** Data mining is field within computing that attempts to discover patterns across large volumes of data. Just as with software engineering, the process of data mining has a life cycle. Within software engineering, testing represents a process across the entire cycle and is represented by a clear and practical model which is developed at the industrial level. For Business Intelligence (BI) projects this research dealing with the validity of the models that are generated via data mining but there isn't a model that exists over its entire life cycle. This paper proposes a test model for BI projects referencing test practices found in software engineering.

**Keywords--** *BI Projects, Model, Testing, Software Engineering*

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2016.1.1.1.271>  
ISBN: 978-0-9822896-9-3  
ISSN: 2414-6390

# Modelo de pruebas en proyectos BI

Katerine Villamizar Suaza, Magister en Ingeniería de Sistemas<sup>1</sup>, Dario Enrique Soto Durán, Magister Ciencias de la Computación<sup>2</sup>, Juan Camilo Giraldo Mejía, Magister en Ingeniería de Sistemas<sup>3</sup>, Jovani Alberto Jiménez Builes, Doctor en Ingeniería de Sistemas<sup>4</sup>

<sup>1</sup>Tecnológico de Antioquia, Colombia, kvillamizar@tdea.edu.co; <sup>2</sup>dsoto@tdea.edu.co; <sup>3</sup>jgiraldo1@tdea.edu.co. <sup>4</sup>Universidad Nacional de Colombia, jajimen1@unal.edu.co.

**Resumen—** La minería de datos es un campo de la computación que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Al igual que la ingeniería de software, el proceso de minería de datos tiene un ciclo de vida. Dentro de la ingeniería de software, las pruebas representan un proceso transversal a todo el ciclo, lo cual está representado por un modelo claro y práctico que se desarrolla a nivel industrial. Para proyectos BI (Business Intelligence traducido al español Inteligencia de Negocios) existen investigaciones sobre las validaciones de los modelos que se generan en la minería de datos, pero no un modelo que sea transversal a todo el ciclo del mismo. Este artículo, propone un modelo de pruebas para proyectos BI instanciando las prácticas de pruebas en ingeniería de software.

**Palabras clave—** Proyectos BI, Modelo, Pruebas, ingeniería de software.

## I. INTRODUCCIÓN

La calidad según la ISO 9000 es el grado en el que un conjunto de características cumple con los requisitos, lo que indica la satisfacción del interesado acerca de un producto de software [1]. Para asegurar la calidad de un producto, se usan las pruebas, lo que representa una revisión final de las especificaciones, del diseño y de la codificación [2].

Diferentes investigaciones se enfocan en presentar modelos, prácticas y metodologías para apoyar el proceso de pruebas de los productos de software [3]. Sin embargo, como se especifica, las pruebas de software se enfocan exclusivamente en el ciclo de vida del mismo y no en el ciclo de vida de proyectos de minería de datos [3].

En el ciclo de vida para proyectos BI, se considera una etapa en el que los modelos generados a partir de diversas técnicas deben ser validados [4]. Sin embargo, se presenta el faltante de las pruebas en todo el ciclo de vida, por lo cual, se propone las pruebas de manera transversal que permita verificar y validar todas las partes del ciclo en correspondencia con los requisitos y objetivos previamente propuestos en un proyecto BI.

Este artículo posee la siguiente estructura: inicialmente, se presenta un marco teórico de los conceptos necesarios para entender la problemática presentada; luego, se analizan diversos trabajos desarrollados sobre las pruebas en proyectos de minería de datos; posteriormente, se propone un modelo de pruebas para proyectos BI y finalmente se discuten las

conclusiones y el trabajo futuro que se deriva de esta propuesta.

## II. MARCO TEÓRICO

### A. Pruebas

En el reciente estándar liberado por ISO / IEC 29119, todas las actividades de verificación y validación realizado durante el proceso de software se considera que pertenecen al proceso de pruebas.

Las pruebas de software se define como: “la actividad en que un sistema o un componente se ejecuta bajo condiciones controladas, los resultados son registrados y la evaluación es realizada sobre algún aspecto del sistema o componente” (ISO / IEC / IEEE, 2010).

Las actividades de verificación y validación de software, se derivan del aseguramiento de la calidad, que tienen como objetivo asegurar que el sistema se construye acorde al proceso de desarrollo y que cumple con las necesidades del cliente [5]. Según Boehm, la verificación se refiere a: “*estamos construyendo el producto correcto*”, y la validación se refiere a: “*estamos construyendo el producto adecuado*” [6]. Por lo tanto, la validación permite confirmar que el software es capaz de cumplir con el uso previsto (ISO / IEC / IEEE 2010) y la verificación confirma que el software cumple con los requisitos definidos por el cliente (ISO / IEC / IEEE 2010), ambos conceptos están relacionados con los objetivos del proceso de pruebas.

La validación se divide en validación estática y dinámica. La validación estática, verifica la exactitud del producto de software sin ejecutar el sistema de software, mientras que la validación dinámica ejecuta el sistema de software [6].

El marco de procesos de las pruebas de software según [7], está integrado por los siguientes subprocesos:

- Procesos de Gestión: Conjunto de procesos que permiten gestionar el proyecto de pruebas.
- Procesos de Soporte: Conjunto de procesos que promueven la ejecución de aquellos procesos que incluyen las actividades propias de verificación y validación de productos software.
- Procesos Técnicos: Conjunto de procesos que incluyen las actividades relacionadas directamente con la verificación y la validación de los productos software.

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2016.1.1.271>

ISBN: 978-0-9822896-9-3

ISSN: 2414-6390

**14<sup>th</sup> LACCEI International Multi-Conference for Engineering, Education, and Technology:** “Engineering Innovations for Global Sustainability”, 20-22 July 2016, San José, Costa Rica.

A nivel técnico, entre los aspectos más relevantes del proceso de pruebas, se destacan, el alcance y el tipo de prueba. A continuación se describen:

#### Alcance de las pruebas

El alcance hace referencia a las etapas de las pruebas, las cuales se definen como [8]:

- Pruebas unitarias: El objetivo es probar las unidades más pequeñas del software, el componente o módulo de software.
- Pruebas de componentes: El objetivo de la prueba de componentes es escoger la pieza más pequeña del software a probar, aislarlo del resto del código, y determinar si se comporta exactamente como se espera. Cada componente, se prueba por separado antes de integrarlo en un servicio.
- Pruebas de integración: Se centran en verificar el ensamblaje correcto entre los distintos componentes, una vez que han sido probados unitariamente.
- Pruebas de sistema: Son pruebas de integración del sistema completo. Permiten probar el sistema en su conjunto y con otros sistemas con los que se relaciona, para verificar que las especificaciones funcionales y técnicas se cumplen. Estas pruebas se realizan en un entorno fuera del alcance del desarrollador.
- Pruebas de aceptación: Los usuarios prueban el sistema o aplicación para establecer si está listo para la salida a producción.

#### Tipos de prueba

De acuerdo con las dimensiones de calidad que se desean evaluar, en [9] las pruebas se clasifican como:

- Las pruebas funcionales, que a su vez se dividen en pruebas de requisitos funcionales, pruebas de control de acceso y seguridad, y pruebas de volumen.
- Las pruebas de usabilidad evalúan la facilidad de uso de las interfaces con el usuario.
- Las pruebas de confiabilidad involucra la ejecución de otros tipos de pruebas como: Pruebas de integridad, estructurales y de estrés.
- Las pruebas de rendimiento incluyen: pruebas de “benchmark” (Procedimiento que evalúan el rendimiento), pruebas de carga y pruebas de perfiles.
- Las pruebas de infraestructura se dividen en pruebas de configuración y pruebas de instalación.
- Las pruebas de regresión.

Para la aplicación de los procesos de verificación y validación se tienen las siguientes técnicas [10]:

- Técnicas Estáticas: Es una técnica de evaluación de artefactos del proceso de desarrollo denominada

generalmente revisiones, pretende detectar de forma manual defectos en cualquier producto del proceso de construcción. Los tipos de revisiones que se realizan son:

- Revisiones Informales.
  - Inspecciones.
  - Walkthrough (Revisión guiada).
  - Auditorias.
- Técnicas Dinámicas: Estas técnicas permite evaluar un sistema o uno de sus componentes. Se ejecuta en circunstancias previamente definidas, comparando los resultados obtenidos y los resultados esperados para identificar las fallas en el software.
    - Caja Blanca o estructural: Se enfocan en examinar los detalles procedimentales del código fuente en el contexto lógico del desarrollo.
    - Caja Negra o Funcionales: Estas pruebas se realizan al sistema de software en tiempo de ejecución orientada a evaluar la funcionalidad que debe realizar.

#### *B. Proyectos BI*

Los proyectos de BI, están orientados a la obtención de conocimiento a partir de los datos consolidados que se encuentran en las bodegas de datos, las cuales son repositorios de información que son alimentados por diferentes fuentes de datos operativos o transaccionales [11].

La obtención de conocimiento denominada KDD (Knowledge Discovery in Databases Traducido al español proceso de extracción de conocimiento) está encaminada a obtener patrones potencialmente útiles y fáciles de comprender [12]. Esta actividad se realiza con la aplicación de técnicas de Minería de Datos, las cuales integran bases de datos, estadística, e inteligencia artificial; encontrando relaciones en los datos, y generando modelos o representaciones del conocimiento obtenido.

El KDD se encarga de preparar, y analizar los datos obtenidos, los cuales dan significado a los patrones encontrados [13]. Durante las etapas de obtención de conocimiento, se utilizan técnicas de aprendizaje inteligente que van analizando los datos en forma rápida. Los datos se encuentran almacenados en repositorios conocidos como almacenes de datos, y la utilización de estos depende de técnicas de descubrimiento de conocimiento, que son utilizadas como mecanismos de recuperación de información. La información obtenida es convertida en conocimiento a través de algoritmos de Minería de Datos, que extraen el conocimiento a partir de ciertos parámetros que se configuran en una etapa denominada entrenamiento [11].

Las etapas tradicionales del ciclo de vida para obtención de conocimiento son: el modelado del negocio, la extracción o comprensión de los datos, la preparación de los datos que implica la selección, limpieza y transformación, la consolidación que implica el modelado de la bodega de datos, la explotación de los datos; que se realiza aplicando técnicas de minería de datos para descubrir patrones, y por último la interpretación y evaluación de los modelos obtenidos [14].

### Metodologías para proyectos BI

Las metodologías se plantean como un conjunto de actividades, y de tareas que se siguen para cumplir un propósito. La metodología indica que y como hacer las tareas. La metodología permite desarrollar las etapas del proceso de obtención de conocimiento en forma ordenada y sistémica [14].

#### Metodología CRISP-DM (Cross-Industry Standar Process traducido del inglés Proceso estándar para minería de datos Cross-Industry)

Es un estándar internacional que ha sido creado para el buen desarrollo de proyectos BI [15]. Se soporta en seis etapas. La primera es la comprensión del negocio, la cual relaciona el modelado de la organización. Posteriormente, se realiza la comprensión de los datos, la cual apunta a extraer datos desde las fuentes u orígenes. Luego, la preparación de los datos, a través de un proceso de extracción. Se continúa con el modelado del repositorio, donde se almacenan los datos y la consolidación de los mismos a través de la transformación y carga de datos, luego se realiza la extracción del conocimiento, a esta etapa se le denomina explotación. El conocimiento viene en modelos generados que son revisados en una etapa que se denomina evaluación. Cuando todo es consistente se puede visualizar. Esta es la última de las fases que indica la metodología CRISP [14]. Ver figura 1. Esta metodología interrelaciona las diferentes fases del proceso entre sí, de tal manera que se consolida un proceso iterativo [16]. Aunque se creó pensando en grandes proyectos, su flexibilidad permite adaptarse para proyectos de cualquier tamaño [17]. La metodología define dos documentos en el desarrollo del proyecto; uno denominado el modelo de referencia, y otro la guía del usuario. El modelo de referencia describe fases, tareas y resultados de un proyecto KDD. El documento guía del usuario, contiene los elementos técnicos relacionados con los aspectos detallados en la guía de referencia. Aquí, se tienen en cuenta listas de comprobación a partir de las tareas de cada etapa [18].

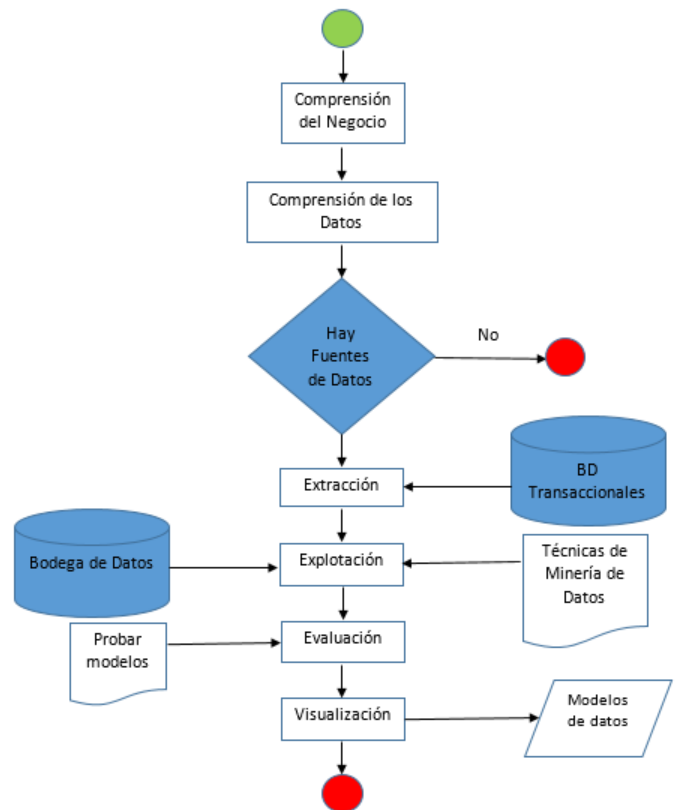


Figura 1 Flujo de Datos Metodología CRISP-DM  
[Elaboración propia]

#### Metodología SEMMA (Acrónimo de: Muestreo, Exploración, Modificación, Modelado y Valoración)

Consiste en seleccionar, explorar y modelar a partir de conjuntos de datos, con el propósito de obtener patrones frecuentes que para la organización son desconocidos. Las fases que involucra esta metodología son: La primera etapa es denominada muestreo, que corresponde a la extracción de los datos, los cuales serán más adelante analizados. La segunda etapa se denomina exploración, en ella se revisan los datos, con el fin de optimizar la eficiencia de los modelos que se obtendrán. Luego se desarrolla la etapa de modificación, en esta, se ajustan los datos para que tenga el formato adecuado y lleguen de manera consistente al modelo analítico. La etapa de modelado, permite definir el modelo de datos analítico donde se consolida la información. Por último, la etapa de valoración consiste en evaluar los modelos obtenidos. Ver figura 2. De esta forma, se garantiza que los datos tengan predicciones confiables [15].

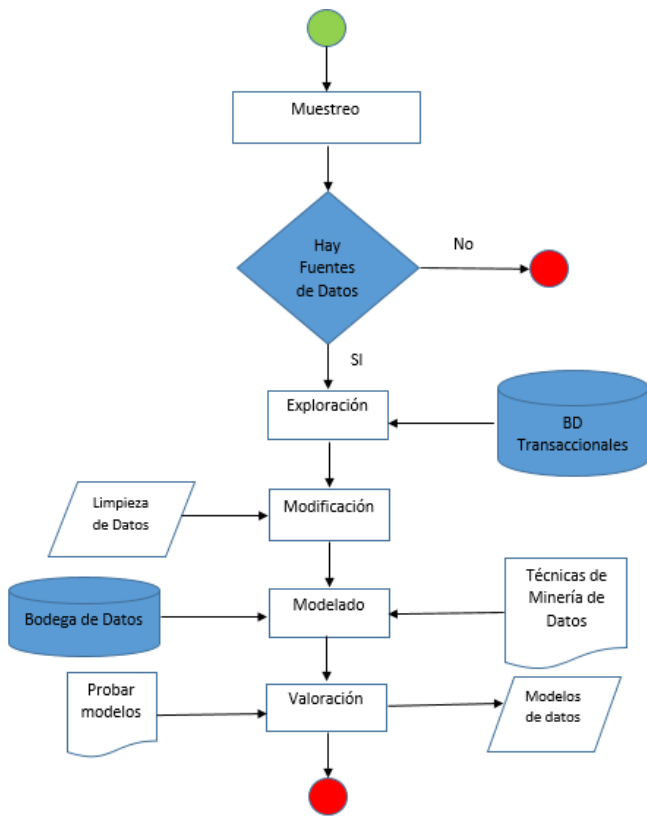


Figura 2 Flujo de Datos Metodología SEMMA [Elaboración propia]

### III. ANTECEDENTES

#### A. Ajuste y Validación

En “introducción a la minería de datos” [19], se plantea posterior a la extracción de datos una etapa de ajuste y validación, tomando como base un conjunto de datos (muestra), en donde se especifica que se deben ajustar los modelos de acuerdo a los datos solicitados. Adicional a esto, se plantea como etapa final la validación del modelo usado, en donde se pone a prueba los datos para identificar patrones para que sean generalizados [19].

Esta validación se realiza en la etapa de extracción y en la etapa posterior a la explotación y anterior a la visualización; sin embargo, no se tiene en cuenta la validación en todas las etapas del proyecto de minería de datos.

#### B. Validación de un modelo de minería de datos

Microsoft, presenta diferentes herramientas para probar y validar modelos de datos con el fin de garantizar el uso adecuado de los diferentes modelos de minería de datos [20].

Sin embargo, estas herramientas hacen parte de la validación del proceso de explotación y no permite la evaluación de todos los procesos del ciclo de vida de proyectos de minería de datos.

#### C. Métricas de evaluación

La matriz de confusión se usa como una medida de evaluación técnica que contiene información acerca de las clasificaciones que se tienen de los datos y las que se predicen para un determinado proyecto, generalmente se basan en una tabla de contingencia para presentar los datos predichos acertados y los datos erróneos, para los cuales se realiza una serie de mediciones técnicas como la precisión, la exactitud, entre otros [21].

Al igual que el planteamiento anterior, esta métrica de evaluación sólo se centra en la parte técnica de los datos y no tiene en cuenta la evaluación de otras etapas de la minería de datos.

#### D. Casos de Validación

En un proceso de transformación de datos para proyectos de explotación de información, se plantea las pruebas de dicha transformación por medio de casos de validación específicos, en donde, se plantea una serie de pasos para evaluar las actividades del proceso y el desempeño de cada uno de los pasos [22]. Esta es una aproximación a una definición de pruebas de verificación en proyectos BI, sin embargo, en este proceso no se plantean las validaciones de manera formal ni se especifica un artefacto o proceso que se deba tener en cuenta a la hora de realizar algún tipo de prueba.

#### E. Herramienta de validación de método de estimación de esfuerzo para proyectos BI

Esta propuesta, presenta diferentes modelos, y un método matemático paramétrico de estimación para proyectos de minería de datos (DMCOMO). En donde, se presenta una alternativa de validación de dicha estimación por medio de una herramienta implementada [23]. La herramienta, podría ser un valor agregado a las validaciones que se pueden hacer y plantear en proyectos de minería de datos, dado que las estimaciones se realizan al principio de la propuesta de un plan de proyecto.

#### F. Hacia la Ingeniería de Data Mining: Un modelo de proceso para el desarrollo de proyectos

Diferentes investigadores, realizan un paralelo entre la ingeniería de software y la ingeniería de Data Mining (DM), teniendo en cuenta que si bien la metodología CRISP- DM de minería de datos, tiene una fase de evaluación de resultados, queda un gran faltante que es la evaluación del proceso en general, en donde se hace necesario, el planteamiento de tareas que permitan evaluar cada una de las fases del ciclo de vida de un proyecto BI. El modelo de proceso planteado, se presenta en la Figura 3, en donde, el proceso de evaluación hace parte de un proceso integral para todo el ciclo de vida de proyectos DM [24]. Esta propuesta, es de gran utilidad a la hora de definir las pruebas en proyectos BI. Sin embargo, no se plantea las actividades ni artefactos que se deben realizar a la hora de ejecutar dicha evaluación.

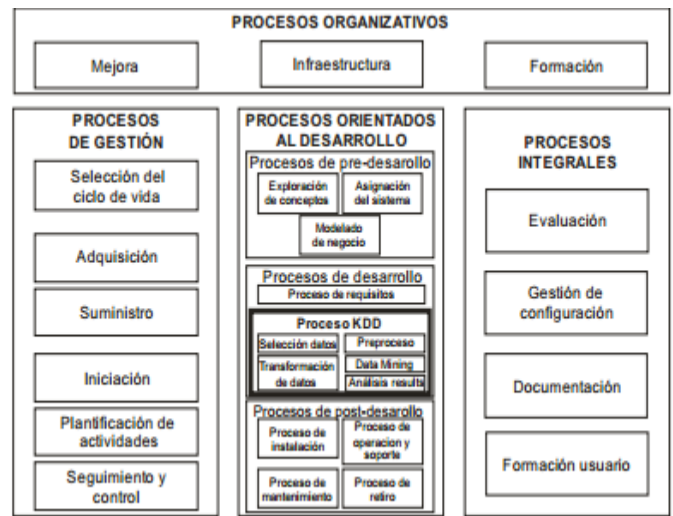


Figura 3 Modelo de procesos para ingeniería de DM [24].

G. Aseguramiento de la calidad para proyectos de explotación de información

Se presenta una propuesta aún no finalizada, en donde se espera dar una guía básica para un proceso de aseguramiento de la calidad en el marco de proyectos de explotación de información. Este artículo es preciso a la hora de definir las pruebas en proyectos BI, sin embargo no se presenta los artefactos ni el detalle del proceso para especificar las pruebas en este tipo de proyectos [25].

Las anteriores investigaciones expuestas, se enfocan en validaciones específicas de una parte del ciclo de vida de proyectos BI. Sin embargo, no se enfocan en determinar que artefactos, ni cómo se deben realizar las pruebas en todo el proceso de los proyectos de minería de datos.

IV. METODOLOGÍA

Como solución a la carencia de un modelo de pruebas para proyectos BI, se referencia el modelo de procesos de desarrollo denominado “Modelo V”, realizando una integración con las etapas del ciclo de vida de proyectos BI. A continuación, se presenta una breve explicación de las etapas BI y definición del Modelo V para pruebas de software.

A. Especificación de etapas del ciclo de vida para proyectos BI

En la Tabla 1, se presentan las diferentes etapas del ciclo de vida en proyectos BI, con sus correspondientes actividades a realizar y los entregables que se generan a partir de dichas actividades. Esto se realiza teniendo en cuenta la especificación de etapas que indica la metodología CRISP-DM.

Tabla 1 Especificación del ciclo de vida en proyectos BI, a partir de la metodología CRISP-DM [14].

ETAPA	ACTIVIDADES	ENTREGABLES
<b>Comprensión del negocio (modelado de la organización):</b> Se realiza la Identificación de la organización y sus procesos. Especificación de necesidades entorno a las metas de la organización.	<ul style="list-style-type: none"><li>• Caracterizar la organización.</li><li>• Realizar una especificación de procesos y actividades relacionados con las necesidades identificadas previamente.</li><li>• Identificación del problema, y alcance de mismo.</li><li>• Realizar preguntas que permiten especificar las necesidades o alcance esperado.</li><li>• Definir los Indicadores articulados con las necesidades y metas esperadas por la organización, y que responden a las preguntas indicadas.</li></ul>	<ul style="list-style-type: none"><li>• Documento con características de la empresa.</li><li>• Diagrama BPMN (notación del modelado de procesos de negocio).</li><li>• Especificación de las variables (hechos, y dimensiones)</li><li>• Listado de preguntas relacionadas con los requisitos o necesidades.</li><li>• Listado de indicadores.</li></ul>
<b>Comprensión de los datos (Extracción,</b>	<ul style="list-style-type: none"><li>• Identificar las fuentes de datos internas y</li></ul>	<ul style="list-style-type: none"><li>• Listado de las bases de datos operativas y</li></ul>

<b>Fuentes):</b> Se obtiene de las diferentes fuentes u orígenes (Bases de Datos) de la organización; las colecciones de datos necesarias para dar respuesta a las necesidades planteadas.	externas (Bases de Datos y sus formatos). • Especificar el mapeo necesario para obtener los datos desde las diferentes fuentes. • Definir las estrategias de extracción de datos, relacionando variables, y criterios de búsqueda. • Realizar de limpieza a las diferentes colecciones obtenidas desde las fuentes internas y externas.	transaccionales internas y externas que contienen los datos necesarios para el análisis. • Formatos específicos compuestos por variables (atributos o campos), para recibir los datos desde los diferentes orígenes. • Scripts (consultas o vistas), que apuntan a los diferentes orígenes, y extraen los datos específicos. • Nuevos conjuntos o colecciones de datos.
<b>Modelado (Consolidación)</b> Se consolidan los datos obtenidos de las diferentes fuentes, en una base de datos denominada bodega de datos.	• Se diseña e implementa la bodega de datos (Modelo estrella), donde se especifican las dimensiones y los hechos.  • Se especifican procedimientos necesarios para migrar los datos desde el origen al destino (desde las fuentes, a la bodega de datos).  • Durante este proceso se realiza una actividad que consiste en transformar los datos (tipos y longitudes), con el fin de	• Diseño de bodega de datos. • Diccionario de la bodega.

	garantizar la simetría y consistencia.	
<b>Explotación</b> Se realizan consultas a la bodega de datos, con el fin de obtener información, y después de analizarla se obtiene conocimiento.	• Se determinan los criterios para definir las colecciones de información desde la bodega de datos.  • Se especifican criterios de selección para las técnicas de análisis de información apropiadas.  • Validación de modelos.	• Consultas (scripts) con condiciones para obtener la información de la bodega.  • Entrenamiento de las técnicas para obtención de conocimiento (parámetros de configuración)  • Generación de modelos de datos en forma automática.
<b>Visualización</b>	• Se especifican las interfaces, para visualización de los modelos obtenidos.	• Datos a través de cuadros de mando, tableros de control, aplicaciones web o móvil.

### B. Modelo V

El modelo en V, es una representación del flujo de actividades o ciclo de vida de un producto de software, se presenta en forma de uve (V) (Figura 4), en donde se realiza un paralelo de las fases de construcción de un producto de software y su correspondencia con las fases de pruebas [26].

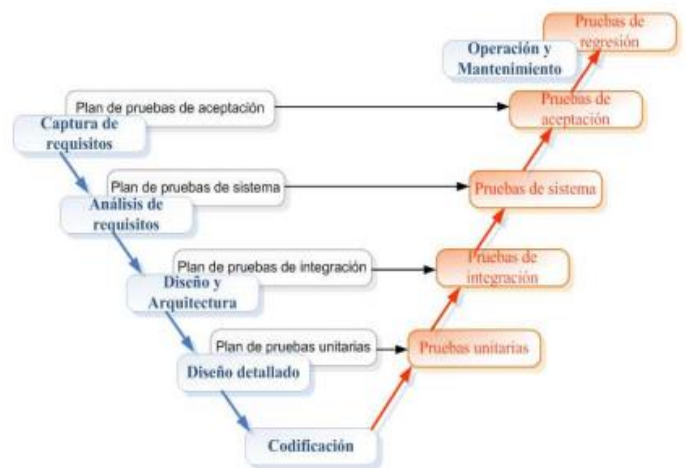


Figura 4 Modelo V - Fases de prueba en un proceso de desarrollo de software [27].

## V. RESULTADO: MODELO PROPUESTO PARA PRUEBAS BI

Como solución a la problemática planteada, a continuación se presenta el Modelo propuesto y la descripción de las pruebas que se deberían realizar en cada una de las etapas del ciclo de vida de proyectos BI. Ver figura 5.

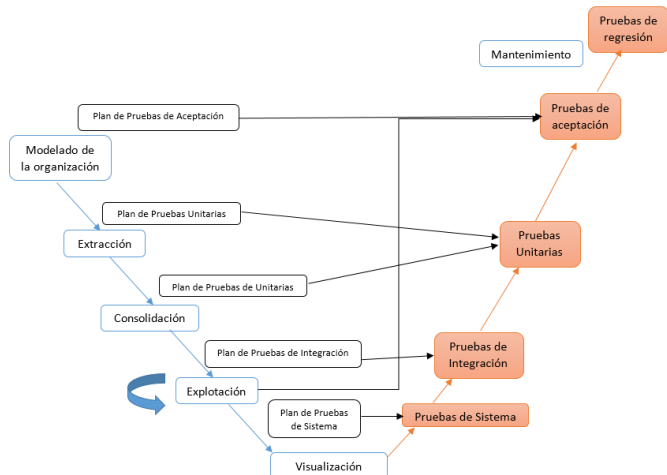


Figura 5 Modelo de pruebas BI [*Elaboración propia*]

- *Modelado de la organización*

En esta fase se debe contrastar las necesidades de los clientes frente al plan estratégico de la organización validando la consistencia de las reglas de negocio y la estructura organizacional [28]. Lo que llevaría finalmente, a la validación por parte del cliente (pruebas de aceptación) cuando finalice el proyecto BI, para corroborar el cumplimiento de los objetivos planteados al iniciar el proyecto.

Para pasar a la etapa de pruebas de aceptación, en esta fase se debe verificar lo siguiente:

- Se debe evaluar la calidad de las fuentes de datos para realizar una revisión temprana del acceso y la estructura con el propósito de evitar reprocesos en la construcción del conjunto de datos [29].
- El conocimiento del cliente frente a las necesidades del proyecto.

Adicional a esto, es importante establecer los criterios de calidad que determinaran el éxito del proyecto en el plano organizacional y en el plano técnico. Para este propósito, se debe establecer un mecanismo que permita valorar ¿Cómo el nuevo conocimiento a implementar generara impacto en los procesos organizacionales o bien en la toma de decisiones estratégicas?

- *Extracción*

En esta fase, el equipo de pruebas debe establecer un instrumento que permita verificar el cumplimiento de la realización de un análisis exploratorio y descriptivo de los datos, con el propósito de familiarizarse con los mismos. En esta misma línea se debe identificar la distribución y comportamiento de las variables con el objeto de proyectar las tareas de limpieza a desarrollar.

El propósito de la verificación es prevenir defectos en los procesos siguientes, en consecuencia, es necesario revisar e inspeccionar la calidad de los datos asociados a las fuentes de información, por medio de pruebas unitarias, dado que las bases de datos transaccionales generalmente no son perfectas y completas. También podría ocurrir la existencia de valores atípicos que influyan en los fallos del estudio. Es esencial la depuración de los datos, para evitar la ocurrencia del fenómeno GIGO (Garbage in, garbage out Traducido del inglés basura que entra y basura que sale. En el contexto de la explotación de la información, el término se emplea para explicar la existencia de patrones de análisis de mala calidad lo cual se debe a la entrada de datos de mala calidad).

Es importante, establecer instrumentos que permitan la validación y verificación de las actividades de transformación de variables y creación de atributos derivados para evitar las inconsistencias en la vista minable.

Una vez cumplido el proceso de limpieza de los datos y las transformaciones, se debe garantizar la consistencia de la información. Por tal razón, se propone la ejecución de un análisis descriptivo final del conjunto de datos finalizados con el propósito de tener una modelización e interpretación depurada.

En el proceso de verificación, se deben establecer casos de prueba derivados de las reglas de negocio para que se permita validar la consistencia de los datos que integra la vista minable y así tener una representación fiable de la realidad de la organización.

- *Consolidación*

Es importante revisar las estrategias de migración (transformación y carga de datos), para garantizar la consistencia e integridad de los datos. Se debe verificar que la estructura de la bodega de datos cumpla con uno de los modelos indicados para ser un proyecto de obtención de conocimiento, a partir del tipo de relaciones, las claves primarias y foráneas. Con lo anterior, se garantiza, la simetría entre las características de los atributos de las tablas que correspondan con los campos de las fuentes de datos, y la estructura de la bodega propuesta. Estos elementos deben ser probados teniendo en cuenta las pruebas unitarias, que garanticen la verificación de las unidades más pequeñas del

ETL (Extract, Transform and Load, Traducido del inglés, proceso de extracción, transformación y carga de datos). Adicional a esto, se debe verificar la consistencia de los datos fuente con los de la bodega de datos construida.

- *Explotación*

Para tener un producto exitoso, se debe efectuar una selección adecuada de las técnicas de minería a incorporar en el proceso, teniendo claro el enfoque a utilizar (clasificación o agrupamiento). Es relevante en esta fase, analizar factores que puedan influir en la selección de las técnicas como: tamaño del conjunto de datos, la naturaleza de las variables entre otros.

Es importante implementar estrategias que permitan realizar una selección adecuada, soportados en el juicio de los expertos técnicos alineado a las necesidades del estudio.

Adicional a esto, se debe establecer criterios para evaluar los resultados obtenidos por la aplicación de las técnicas, esta definición se realiza antes de la aplicación de las mismas para evitar sesgos en su validación. En esta misma línea el equipo de pruebas debe realizar la clasificación del conjunto de datos para prueba y entrenamientos de los modelos. Lo que implica pruebas integrales a la hora de ejecutar todo el proceso ETL hasta la aplicación de técnicas en la bodega de datos.

Partiendo de la evaluación inicial de los modelos obtenidos realizada desde el criterio técnico (tasas de error), se debe establecer casos de prueba; los cuales se puedan comprobar en la siguiente etapa, y así, determinar cuales se ajusta a los objetivos propuestos por el proyecto.

En este punto las pruebas deben garantizar la alineación de los modelos obtenidos con los objetivos del negocio, analizando, si los resultados generan valor al propósito organizacional. En donde, es importante involucrar a los usuarios para validar los patrones obtenidos (Pruebas de aceptación).

Una vez analizados los modelos de forma individual, es relevante valorarlos, en función de los que mejor se ajustan a los objetivos planteados a nivel técnico y organizacional. En esta fase, se debe verificar el planteamiento de caminos alternos, en caso que no se obtenga resultados exitosos en el descubrimiento de patrones para el conjunto de datos; en caso tal de que se haya seleccionado el modelo incorrecto, se vuelva a iniciar esta etapa con la selección de una nueva técnica.

- *Visualización.*

Hasta este punto, los objetivos técnicos del estudio se han obtenido, pero los objetivos de negocio solo se garantizarán, si los usuarios de la solución apropian de forma adecuada el

nuevo conocimiento, a la toma de decisiones y a la generación de fortalecimiento de los procesos organizacionales, que en algunos casos se refleja en la ampliación de los objetivos estratégicos con el propósito de conquistar nuevos mercados. Por tal razón, las pruebas de software deben verificar el cumplimiento de un plan de implantación a producción, donde se pueda comprobar los perfiles requeridos tanto del capacitado como del formador. Al igual que se deben realizar pruebas de sistema como un todo, corroborando, desde los procesos hasta la visualización como parte de los objetivos iniciales del proyecto BI, con el fin de satisfacer, las necesidades propuestas alrededor de diversos indicadores propuestos.

- *Mantenimiento.*

Cómo todo proyecto, posterior a la implantación en producción (una vez implementados los modelos), el método de pruebas, debe establecer criterios de seguimiento que permita garantizar la validez de los modelos frente a los cambios del entorno, que se pueden presentar en las variables organizacionales. Es aquí, donde se deben realizar pruebas de regresión a la hora de implementar cambios en todo el proyecto, con la finalidad de corroborar que todo el proceso del proyecto se sigue ejecutando con calidad, sin tener que modificar el *core (núcleo)* del mismo.

## VI. CONCLUSIONES

Es claro que las soluciones BI siguen una secuencia de etapas que soportan el proceso de construcción; al igual que los procesos de ingeniería de software. En consecuencia, es relevante plantear prácticas de verificación y validación que permita probar la calidad y el éxito del proyecto.

A nivel de las actividades de verificación que integra las pruebas, es necesario establecer instrumentos y mecanismos permanentes que permitan analizar la completitud y consistencia de la construcción de la solución BI.

Un elemento esencial en la construcción de una solución BI, son los activos de proceso derivados de cada etapa de construcción, porque son de gran utilidad para la implementación y evolución del sistema.

Adicional a esto, es importante incluir las buenas prácticas identificadas en la gestión de proyectos, como las lecciones aprendidas que son claves para evitar la presencia de errores en los nuevos proyectos. Para el modelo propuesto, se plantea un análisis post-mortem al final de cada etapa; con el propósito de identificar elementos claves para el inicio de la siguiente fase o en caso de ser necesario el retorno a la fase anterior para los ajustes requeridos de la solución.

## VII. TRABAJO FUTURO

Como trabajo futuro, se pretende definir un marco de procesos, roles y artefactos que permitan guiar el desarrollo de pruebas BI en el contexto de los proyectos de minería de datos con base en el modelo propuesto.

En consecuencia a lo anterior, se propone el desarrollo de una herramienta de gestión de pruebas para proyectos BI, que garantice el cumplimiento de los elementos planteados en cada etapa de pruebas para proyectos de minería de datos.

Presentar una propuesta de estimación de pruebas BI, realizando una extensión al modelo estimación DMCOOMO, asociado al marco de procesos de pruebas propuesto.

## VIII. REFERENCIAS

- [1] ISO. Norma internacional ISO 9000 - Sistemas de gestión de la calidad - Fundamentos y vocabulario. Ginebra: ISO 9000. 2005.
- [2] Alonso, F. Martínez, L. Segovia F. Introducción a la ingeniería del software Modelos de desarrollo de programas. España: Delta. 200.
- [3] Glenford, J. Wiley, J. The art of software testing. 2004.
- [4] Bravo, C. Madonado, S. Weber, R. *Experiencias prácticas en la medición de microempresarios utilizando modelos de credit scoring*. Revista ingeniería de sistemas. Vol XXIV. 2010.
- [5] Wallace D. R. and R. U. Fujii. Software verification and validation: an overview, IEEE Software, Vol. 6, No. 3, pp. 10-17, May 1989.
- [6] Boehm B.W. Software engineering economics, IEEE Trans. on Software Eng., Vol. 10, no. 1, Jan. 1984. pp. 4 - 21.
- [7] Sanz, A., García, J., Saldaña, J. & De Amescua, A. A proposal of a process model to create a Test Factory, International Conference on Software Engineering (ICSE 2009 - Workshop of Software Quality). Vancouver. Canadá. 2009.
- [8] Baker, P., Ru Dai, Z., Grabowski, J., Haugen, O., Schieferdecker, I., & Williams, C. Model Driven Testing – Using the UML Testing Profile, Springer. 2008.
- [9] Shuja, A., K., & Krebs, J. IBM Rational Unified Process Reference and Certification Guide: Solution Designer, IBM Press. 2007.
- [10] JURISTO Natalia, MORENO Ana M., VEGAS Sira. Técnicas de Evaluación de Software. 131 p. / UNIVERSIDAD SIMON BOLIVAR. Prueba repetible y mantenible. [en línea]. Bitácora de Prueba. Ingeniería de Software 3. Enero-Marzo 2008. [en línea]. [consultado 23/02/2016]. Disponible: <http://www.tuobra.unam.mx/publicadas/040803214240.html>.
- [11] Lezcano, Ramón David. “Minería de Datos”. Trabajo de Investigación. Universidad Nacional del Nordeste. 2002.
- [12] Cabrera Rodríguez, Julio A. “Bodega de Datos”. Profesor Asistente Grupo de Técnicas de Dirección. Universidad Agraria de la Habana. 1998.
- [13] Violeta Valcárcel Asencios. “Data Mining y el descubrimiento de conocimiento”. Revista de la Facultad de Ingeniería Industrial. Vol. (7) 2: pp. 83-86. UNMSM ISSN: 1560- 9146. Diciembre de 2004.
- [14] Handbook of Software Engineering and Knowledge Engineering Fundamentals, World Scientific Publishing Co., Singapore, pp 615-637, Jan 2001.
- [15] Paola Britos. Objetivos de Negocio y Procesos de Minería de Datos en Sistemas Inteligentes. Reportes Técnicos en Ingeniería del Software. 7(1): 26-29 ISSN 1667-5002.
- [16] Salcedo Parra, Octavio j.; Galeano, Rita Milena; Rodríguez B., Luis G. Metodología Crisp para la Implementación Data Warehouse. Tecnura, vol. 14, núm. 26, enero-junio, 2010, pp. 35-48 Universidad Distrital Francisco José de Caldas Bogotá, Colombia.
- [17] E. Fernández, H. Merlino, M Ochoa, E. Diez, P.Britos, R García. Gestión Asistida de Documentos en una Metodología de Explotación de Información. II Workshop de Ingeniería del Software y Bases de Datos XI Congreso Argentino de Ciencias de la Computación. Pág. 504-514. 2005.
- [18] Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información. CACIC 2010 - xvi congreso argentino de ciencias de la computación.
- [19] L. Vieira Braga., L. Ortiz Valencia. “Introducción a la minería de datos.” e-papers. Rio de Janeiro. 2009.
- [20] Microsoft. “Proyectos de minería de datos. Probar y validar modelos”. Developer Network. 2014.
- [21] F. Barrientos, S. Ríos. “Aplicación de Minería de datos para predecir fuga de clientes en la industria de las telecomunicaciones.” Revista ingeniería de Sistemas. 2013.
- [22] R. García Martínez, D. Rodríguez. “Un proceso de transformación de datos para proyectos de explotación de información”. Trabajo Final. Universidad Nacional de Lanus. 2013.
- [23] P. Pytel. “Método de estimación de esfuerzo para proyectos de explotación de información. Herramienta para su validación.” Tesis de maestría. Universidad politécnica de Madrid instituto tecnológico de buenos aires. 2011.
- [24] G. Mariscal. O. Marbán. A. González, J. Segovia. “Hacia la Ingeniería de Data Mining: Un modelo de proceso para el desarrollo de proyectos”. Departamento de Lenguajes y Sistemas Informáticos e Ingeniería de Software. II Congreso Español de informática. 2007.
- [25] Diez, E., Pytel, P., Rodríguez, D., García, R., Lacabanne, M., Leonardis, L., Martins, S., Cartanilica, A., García-Martínez, R. “Aseguramiento de la calidad para proyectos de explotación de información”. XIV Workshop de investigadores en Ciencias de la computación. 2012.
- [26] L. Shuping and P. Ling. The research of v model in testing embedded software. pages 463–466, 29 2008-Sept. 2 2008.
- [27] A. Torres; “Método de pruebas de Sistema basado en modelos navegacionales en un context MDWE”. Lenguajes y Sistemas informáticos. Universidad de Sevilla. España. 2009.
- [28] J. Montilva C; J. Barrios A. “BMM: A business modeling method for information systems development”. Universidad de los Andes. Facultad de ingeniería. Colombia. 2012
- [29] N. Dib, D. “Evaluación de calidad de datos estadísticos”. IB Revista virtual. Revista de la información básica. Colombia. 2007.