

Adecuación de una Propuesta Metodológica de Enfoque “Híbrido” para la Gestión de Proyectos de Ciencia de Datos

Patricia Cristaldo¹, Esteban Schab¹, Cristhian Richard¹, Ramiro Rivera¹, Anabella De Battista¹, Soledad Retamar¹, Norma Herrera²

(1) *Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Departamento de Ingeniería en Sistemas de Información*

(2) *Universidad Nacional de San Luis, Departamento de Informática*

Resumen

La aplicación de metodologías de gestión de proyectos conlleva a contribuir con los factores de éxito en la gestión de los mismos. Esta propuesta se deriva de un trabajo anterior y considera los aspectos relevantes de la gestión tradicional y ágil, permitiendo la alineación de los objetivos de un proyecto de ciencia de datos con los resultados del mismo. Este trabajo presenta la adecuación y posterior validación de una propuesta metodológica de enfoque “híbrido” a un proyecto de ciencia de datos, considerando el cumplimiento de los factores de éxito afectados a este tipo de proyectos.

1. Introducción

Desde hace un tiempo, en el ámbito de la gestión de proyectos, las metodologías de gestión de proyectos han adquirido una generalizada popularidad, tanto en organizaciones públicas como privadas. Estas metodologías de gestión de proyectos, según su filosofía de desarrollo se pueden clasificar en dos enfoques: *tradicionales*, que se basan en una fuerte planificación durante todo el desarrollo y, *ágiles*, en las que el desarrollo es incremental, cooperativo, sencillo y adaptado. En la industria de la minería de datos se viene generando una creciente conciencia de las partes interesadas en utilizar metodologías que guíen a los proyectos hacia la gestión exitosa [1]. La minería de datos se puede clasificar en diferentes perspectivas: desde el aspecto de los datos, se pueden abordar sistemas de clasificación supervisados, semi-supervisados o no supervisados; desde el punto de vista de las tareas propias de la minería de datos, que pueden ser descriptivas (descubrir patrones o relaciones describiendo los datos) o predictivas (clasificar nuevos

datos basándose en conjuntos de datos disponibles); y desde el ángulo del algoritmo aplicado, que puede ser clasificación, predicción, agrupamiento, asociaciones, entre otros. Dado que todas las prácticas mencionadas anteriormente deben llevarse a cabo siguiendo ciertos pasos guiados, para ser manejable y trazable una colección de procesos de minería de datos [2] han sido propuestas primeramente metodologías como CRISP-DM [3] y SEMMA [4] ambas de enfoque tradicional.

Es muy importante entender que para llevar adelante un proyecto de ciencia de datos, antes de comenzar a ejecutar algoritmos para descubrir patrones interesantes de datos, es necesario gestionar el proyecto correctamente para no producir patrones insignificantes o ininteligibles, que eventualmente pueden conducir al fracaso del proyecto. Es decir, la gestión correcta del proyecto debe lograr una comprensión válida de los datos y el descubrimiento novedoso de patrones. Además, siguiendo un marco de trabajo común en la gestión de proyectos de minería de datos, podría ayudar como una hoja de ruta mensurable para todos los interesados del proyecto.

Sin embargo, la gestión de proyectos sigue siendo un esfuerzo altamente problemático [5], y siguen mostrando exiguos niveles de éxito [6, 7] debido a varias razones [8, 9, 10, 11, 12, 13]. Sin embargo en la última década, se reconoce que las metodologías de gestión de proyectos tienen un impacto significativo en la gestión de los mismos [14]. Es decir, el objetivo de aplicar metodologías o guías de buenas prácticas a los proyectos implica proporcionar una solución general y disminuir varias razones de fracaso en la gestión de los proyectos.

En estudios recientes se examina que la mayoría de los proyectos de software utilizan una combinación de los métodos tradicionales y ágiles [15, 16, 17]. En un trabajo anterior [15] se ha propuesto una metodología de gestión de proyectos de TICs en un ámbito

particular, logrando buenos resultados a partir de la fusión entre las metodologías SCRUM [18, 19] y PRINCE2 [20] y las guías de buenas prácticas ATERN¹ y PMBOK [21].

En el presente trabajo se propone abordar la adecuación de la propuesta metodológica de enfoque “híbrido” para la gestión de proyectos de TICs, en la gestión de proyectos de minería de datos, permitiendo así contribuir a mejorar la gestión de este tipo de proyectos.

El resto de este documento está organizado de la siguiente manera. La Sección 2 analiza los trabajos relacionados a la luz de las distintas variantes que proponen, particularmente enfocándose en la fusión de metodologías de enfoque tradicional y metodologías de enfoque ágil. En la Sección 3 se describen los fundamentos a partir de los factores críticos de éxito en la gestión de proyectos de minería de datos y en la gestión de proyectos en el enfoque ágil. La Sección 4 describe la nueva propuesta metodológica para la gestión de proyectos de minería de datos. La Sección 5 puntualiza la validación de dicha propuesta a un caso de estudio real. Finalmente, en la Sección 6 se presentan las conclusiones y trabajos futuros

2. Trabajos Relacionados

En la literatura existen varias propuestas de metodologías de gestión de proyectos y particularmente de gestión de proyectos vinculados con la minería de datos.

La metodología ágil para el desarrollo de proyectos de minería de datos, SCRUM-DM [22] consiste en la asociación de la metodología ágil SCRUM con la metodología tradicional Crisp-DM. SCRUM-DM proporciona un marco donde el desarrollo del proyecto de data mining sigue los lineamientos de Crisp-DM y la gestión del trabajo sigue el esquema de la metodología SCRUM. El proyecto se inicia a partir de la fase de *Comprensión del Negocio* de Crisp-DM donde se realiza el análisis de los objetivos y del negocio y finaliza en la fase de *Implantación* donde se realiza la integración de los resultados de la minería de datos. El desarrollo se realiza en la fase intermedia, de *Sprint*, donde se contemplan las demás fases de Crisp-DM.

Otra metodología propuesta es ASD-DM (Adaptative Software Development- Data Mining) [23]. Es una solución basada en principios ágiles para aplicaciones predictivas de minería de datos. Se basa en la metodología ASD propuesta originalmente por el mismo autor, basada fuertemente en un principio iterativo de prueba y error. Este marco asegura continuo aprendizaje e intensa colaboración entre

desarrolladores, testers y clientes de minería de datos, reduciendo el tiempo de desarrollo en proyectos predictivos de minería de datos.

Otro trabajo propuesto es [24], que consiste en una adaptación del KDD clásico que toma conceptos de CRISP-DM y del manifiesto ágil. Es una propuesta que se centra en establecer metas iterativas y ciclos más cortos entre los entregables. Los autores en este trabajo establece los puntos en común entre las metodologías de minería de datos y el manifiesto ágil.

En [25] se analiza el proceso de minería de datos y los desafíos técnicos que se enfrenta un científico de datos durante el proyecto, logrando un ciclo de trabajo iterativo muy similar al descrito en el manifiesto ágil.

Durante los últimos años IBM [26], ha sido una de los grandes promotores de la metodología CRISP-DM, y en el año 2015 realizaron, junto con otras empresas participantes del proyecto, una revisión de la metodología CRISP-DM obteniendo como resultado a la metodología ASUM-DM, la cual incorpora principios ágiles, extendiendo y refinando la metodología CRISP. ASUM-DM tiene un enfoque hacia el sector empresarial.

En los últimos años varios autores han propuesto metodologías alternativas a CRISP-DM, no obstante la mayoría difiere muy poco, limitándose a una mera reinterpretación de algunas de las fases que componen la misma con el fin de facilitar su adaptación a un campo u organización particular, a saber: la adecuación de la metodología SEMMA para gestionar un proyecto ligado al sector bancario y de inversiones y con un fuerte énfasis en la fase de implantación [27]; la adaptación de CRISP-DM a la disciplina de la medicina, proponiendo una nueva descomposición de la metodología [28].

Por otro lado autores como Mariscal [29], Marchx [30] y Lopuszynski [31] han realizado comparaciones entre la filosofía planteada en la literatura relacionada a las metodologías tradicionales de minería de datos y los principios del manifiesto ágil, concluyendo en que las metodologías tradicionales se beneficiarían ampliamente con la adopción de los principios ágiles.

3. Fundamentos para la Adecuación de la Propuesta Metodológica de Enfoque “híbrido” a la gestión de proyectos de ciencia de datos

Como se indicó anteriormente, el marco de trabajo de la propuesta metodológica para gestionar proyectos de minería de datos debe poseer un andamiaje construido sobre dos pilares. La subsección 3.1 discute el primer pilar, es decir, los factores críticos de éxito de los proyectos que se gestionan en el enfoque ágil. La

¹ <http://www.dsdm.org/dig-deeper/book/dsdm-atern-handbook>

Subsección 3.2, analiza el segundo pilar, el que se corresponde a los factores críticos de éxito en los proyectos de minería de datos. Por último, en la Subsección 3.3, se determina un conjunto de factores estratégicos integrados para el propósito de la evaluación y mejora de la propuesta metodológica para la gestión de proyectos de minería de datos.

3.1. Factores críticos de éxito para la gestión de proyectos de minería de datos

Los factores críticos de éxito en la gestión de los proyectos son elementos vitales para que un proyecto

tenga éxito. Por ello es necesario especificarlos claramente, luego convertirlos en requerimientos claves que cuantifiquen los objetivos de gestión de proyectos establecidos y acordados, y que permitan medir el desempeño del mismo [32]. Sobre la numerosa bibliografía en la temática, en la Tabla 1 se recopilan los factores críticos de éxito claves en la gestión de proyectos de minería de datos. Estos factores propuestos por algunos autores son resultados de la experiencia en base a encuestas realizadas a profesionales para la gestión de proyectos de minería de datos [23, 24, 28].

Tabla 1. Factores de éxito en la gestión de proyectos de minería de datos

Dimensión	Factores de éxito
Relaciones Personales	Colocar en primer lugar los requerimientos del cliente
	El cliente debe estar comprometido con el proyecto
	Comunicación constante entre el cliente y el equipo de trabajo
	Comunicación dentro del equipo de trabajo
Planificación	Establecer objetivos claros antes de comenzar el proyecto
	Selección adecuada del equipo de trabajo
	Planificación adecuada del proyecto
	Definición correcta del plazos y costos
Minería y análisis de Datos	Selección adecuada de los datos
	Entendimiento adecuado del área de negocio
	Uso de tecnología adecuada
	Interpretación correcta de los resultados
Implantación	Feedback constante
	Asegurar la privacidad de los datos
	Implementación por pasos
Control/ Monitorización/ Entrenamiento	Los resultados pueden no ser inmediatos
	Asegurar la privacidad de los datos

3.2. Factores críticos de éxito para la gestión de proyectos en el enfoque ágil

La Guía PMBOK en su quinta edición (2013), pone énfasis en el papel de los interesados en el proyecto y la gestión de sus expectativas como factores de éxito de proyectos TICs [33].

Lock (2007) [34] establece que del conjunto de factores críticos de éxito de un proyecto, los más importantes a tener en cuenta son los vinculados a las habilidades en la administración de personas. Entre los cuales cita: el bienestar, la seguridad, la motivación de los miembros del equipo, la resolución rápida de conflictos y la adecuada y frecuente comunicación entre los miembros del equipo que lleva adelante el proyecto.

Por otro lado, Chow y Cao [9], distinguen factores de éxito en los proyectos que se gestan utilizando un enfoque ágil, y los clasifican en cinco dimensiones: organización, personas, procesos, técnicas y proyecto, los cuales se muestran en la Tabla 2. Estos datos, según sus autores, fueron recogidos de un estudio realizado a más de cien proyectos de diversas organizaciones ubicadas en diferentes lugares geográficos. Sus autores concluyen que los factores más relevantes para el éxito de los proyectos son: correcta estrategia de entrega, buen ejercicio de técnicas de ingeniería ágiles y equipo de gestión de alta capacidad.

Tabla 2. Factores de éxito. Enfoque ágil (Chow y Cao, 2008)

Dimensión	Factor de éxito
Organización	<ul style="list-style-type: none"> • Fuerte apoyo ejecutivo. • Administrador o patrocinador comprometido. • Cultura de la organización cooperativa, en lugar de jerárquica. • Cultura de comunicación oral o cara a cara. • Establecimiento de entorno de trabajo de estilo ágil.
Personas	<ul style="list-style-type: none"> • Miembros del equipo competentes y con experiencia. • Miembros del equipo con gran motivación. • Coherencia y trabajo en equipo auto-organizado. • Buenas relaciones con los clientes.
Proceso	<ul style="list-style-type: none"> • Agilidad en la gestión de requerimientos. • Agilidad en la gestión de proyectos. • Fuerte comunicación diaria cara a cara. • Fuerte presencia y compromiso con el cliente. • El cliente tiene plena autoridad.
Técnica	<ul style="list-style-type: none"> • Normas de codificación definidas previamente. • Diseño simple. • Cantidad correcta de documentación. • Adecuada formación técnica del equipo.
Proyecto	<ul style="list-style-type: none"> • Proyecto dinámicos, agenda acelerada. • Proyectos con equipos pequeños. • Evaluación de costos por anticipado. • Evaluación de riesgos por anticipado.

La bibliografía sugiere que la combinación de ambos enfoques, tradicional y ágil, es la solución que han encontrado varias organizaciones para la gestión de sus proyectos [35] e incluso ambos enfoques pueden complementarse en un mismo proyecto [36, 37, 38].

A partir de los datos provistos por ambas tablas, se concluye que los aspectos referidos a las dimensiones: *Minería y Análisis de Datos, Implantación y Control/Monitoreo/Entrenamiento* de la Tabla 1, se pueden incluir en la dimensión *Técnica* de la Tabla 2. De esta forma quedarían bien establecidos los factores críticos de éxito para gestionar proyectos de minería de datos.

Por otra parte, estos factores de éxito nos permiten analizar y comparar las características y restricciones de las metodologías y guías de gestión de proyectos de minería de datos existentes, de enfoque tradicional y ágil.

A partir de las dimensiones propuestas por la bibliografía se han formulado cuatro perspectivas, proporcionando un marco y una estructura que se espera que cumpla una metodología que gestione proyectos de minería de datos. Las perspectivas son:

3.2.1 Formación y Crecimiento: esta perspectiva incluye todos aquellos aspectos relacionados con la gestión de los recursos humanos necesarios para poder ejecutar un proyecto de minería de datos, habilidades,

motivaciones, etc. Los requisitos involucrados en esta perspectiva son los inductores necesarios para el desarrollo de equipos de trabajo.

3.2.2 Procesos internos: esta perspectiva incluye todos aquellos aspectos relacionados con la gestión de los procesos del proyecto, que el jefe del mismo deberá acentuar de manera tal de lograr la satisfacción de los interesados. Las autoridades y el jefe de proyecto, deben identificar los procesos críticos en los que se debe sobresalir con excelencia si es que quieren satisfacer los objetivos que ellos mismos se propongan y los objetivos de los demás interesados en el proyecto.

3.2.3 Interesados: esta perspectiva incluye todos aquellos aspectos vinculados con la satisfacción de los interesados. Se trata de identificar claramente los interesados potenciales y la satisfacción de sus necesidades. Los requisitos involucrados en esta perspectiva representan el grado de apoyo y compromiso de los interesados, los cuales valoran y eligen la propuesta de valor que ellos mismos entregaran al proyecto.

3.2.4 Técnicas: esta perspectiva incluye todos aquellos aspectos relacionados con las técnicas de ingeniería vinculadas con el proyecto, tales como: diseño, documentación, pruebas, tecnología, etc.

La Tabla 3, resume los principales aportes de cada metodología o guía de gestión de proyectos, más utilizadas en el mercado, en relación a los factores críticos de éxito contenidos en las perspectivas consideradas precedentemente. Los factores críticos de éxito en la gestión de proyectos de minería de datos se utilizan para realizar la comparación de características y restricciones de las metodologías y guías de gestión de proyectos. La Tabla 3, muestra que si bien, la mayoría de los factores críticos de éxito son considerados por las metodologías y guías analizadas, los mismos no están

todos presentes en una única metodología de gestión de proyectos. Del análisis realizado, puede concluirse que una metodología no es suficiente y que no debería existir la aplicación de varias metodologías dentro del contexto de la organización ya que conviene que todos los interesados en la gestión de diversos proyectos de minería de datos, estén alineados a una misma metodología. Como consecuencia debido a las limitaciones de los enfoques analizados, la integración de ambos enfoques, tradicional y ágil, a través del desarrollo de un enfoque híbrido.

Tabla 3. Grado de adecuación de los factores de éxito y las metodologías y guías de gestión de proyectos de minería de datos

Dimensión	Factores de éxito	Enfoques			
		Tradicional		Ágil	
		PMBOK	CRISP-DM	SCRUM	ATERN
Formación y Crecimiento	Dinámicas de equipo en la gestión de proyecto	X		X	X
	Compromiso, colaboración y disponibilidad de los miembros del equipo e interesados	X	X		X
	Conductas de apoyo emocional entre los miembros del equipo	X		X	
	Recursos especializados según el rol que ocupa cada miembro del equipo de proyecto	X	X		X
	Gestión eficaz del director del proyecto	X	X		
	Coordinación y competencias de los miembros del equipo	X	X	X	
Procesos Internos	Definición de requerimientos, alcance y limitaciones	X	X		X
	Objetivos claros de negocios	X	X	X	X
	Políticas y procedimientos de gestión de proyectos	X		X	X
	Planificación adecuada	X	X		X
	Prácticas ágiles en los procesos	X	X	X	X
Interesados	Apoyo de la organización	X	X		
	Participación de otros interesados	X	X		
	Igual percepción entre los interesados y los profesionales de SI	X			X
	Cultura de la comunicación	X		X	X
	Definición clara de los interesados y su vinculación con el proyecto	X	X	X	
	Contratistas o patrocinadores comprometidos	X	X	X	
Técnicas	Gestión de la comunicación		X		
	Selección adecuada de los datos		X		
	Interpretación correcta de los resultados		X		
	Seguimiento y control del proyecto	X			
	Gestión de riesgos	X			X
	Gestión de la contratación	X			
	Tecnología y herramientas adecuadas	X			
Documentación	X		X		

4. Adecuación de la Propuesta Metodológica de Enfoque “híbrido” a la gestión de proyectos de ciencia de datos

La nueva propuesta metodológica para gestionar proyectos de minería de datos, surge a raíz de la necesidad de agilizar las metodologías creadas para gestionar estos proyectos. Es decir, el incremento de la velocidad, el volumen y la variedad de los datos disponibles en las organizaciones, ha aumentado de sobremanera. A raíz de lo cual ha crecido el interés y el número de proyectos relacionados con el aprovechamiento de los datos [39]. Esto deriva además, que los diversos interesados en estos proyectos sean un mayor número de profesionales de diferentes disciplinas. Saltz (2015) [39] resalta que las metodologías tradicionales no se han adaptado a las exigencias que presentan la gestión de proyectos de minería de datos, ya que se centran únicamente en la mejora de las técnicas de extracción y análisis de datos.

Con el objetivo de poder dar cumplimiento a los factores críticos de éxito propuestos en la Tabla 3, se plantea a continuación una propuesta metodológica de enfoque “híbrido” para la gestión de proyectos de ciencia de datos. Esta nueva propuesta, se basa en dos metodologías de gestión de proyectos. La primera: CRISP-DM (Cross-Industry Standard Process for Data Mining) [40], cuyo modelo de procesos se muestra en la Figura 1. Es una metodología de enfoque tradicional especialmente desarrollada para proyectos de Data Mining. Actualmente está considerada como un estándar “de facto” en el mundo empresarial. Propone un conjunto de actividades que hay que llevar a cabo en el desarrollo de un proyecto de minería de datos y cada una de las actividades se divide en distintas tareas.

La segunda metodología es, una propuesta para la gestión de proyectos TICs diseñada y probada en otro contexto [15]. Esta propuesta metodológica, que muestra la Figura 2, es de enfoque “híbrido” ya que surge como resultante a partir de la fusión de las siguientes guías de buenas prácticas en gestión de proyectos: PMBOK [21] de enfoque tradicional y ATERN de enfoque ágil y; de las metodologías de gestión de proyectos: PRINCE2 [20] de enfoque tradicional y SCRUM [19] y APM [41] de enfoque ágil. Siguiendo los lineamientos de la guía de buenas prácticas en gestión de proyectos, PMI divide un proyecto en tres fases: iniciación, intermedia y de cierre, por otra parte considera en su PMBOK (2013) cinco grupos de procesos: iniciación, planificación, ejecución, control y cierre. Charvat [42] establece que cada metodología de proyecto debe contener fases, y aunque varíen según el proyecto o la industria, en

general las fases incluyen: concepto, desarrollo, implementación y apoyo.

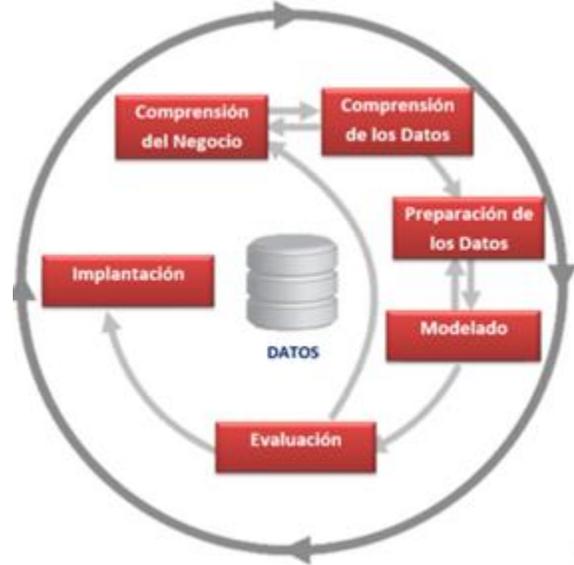


Figura 1. Metodología CRISP-DM

En consecuencia, la metodología propuesta se divide en tres fases: definición y planificación, ejecución y control y, evaluación final y cierre, basados en el ciclo de vida de PMI (2013). El ciclo de vida de la metodología propuesta para la gestión de proyectos de ciencia de datos se muestra en la Figura 3. La misma presenta un desglose de la metodología CRISP-DM en las dos primeras fases de la propuesta metodológica de enfoque “híbrido”, de forma tal de complementar los lineamientos tradicionales propuestos por CRISP-DM con los ágiles, que accede la propuesta metodológica siguiendo fuertemente los pasos de la metodología ágil, SCRUM.

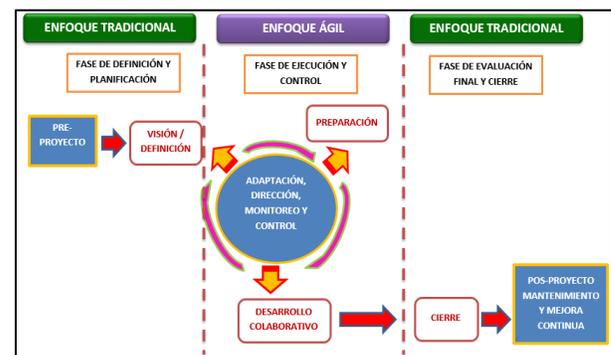


Figura 2. Propuesta Metodológica de enfoque “híbrido”

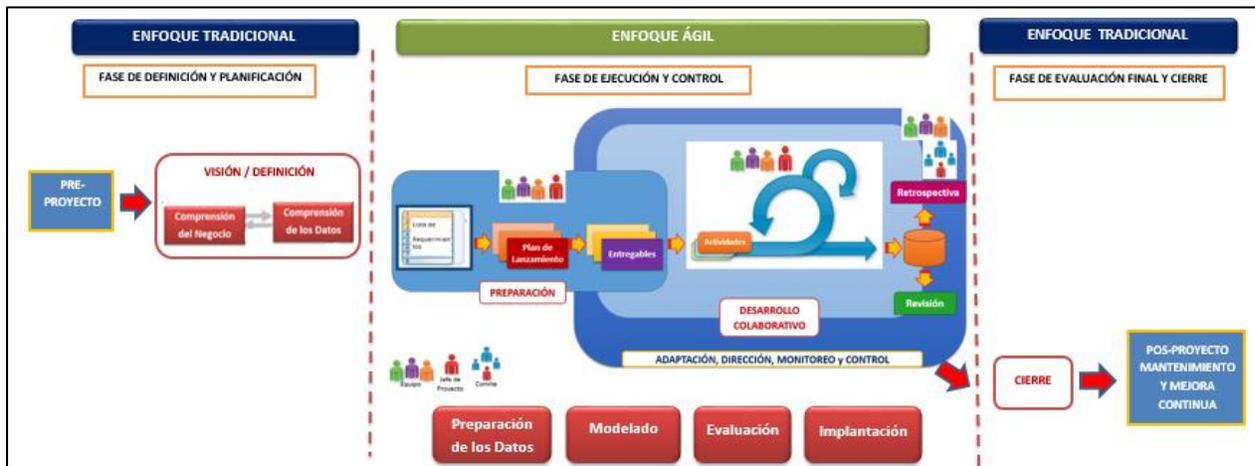


Figura 3. Propuesta Metodológica de enfoque “híbrido” para la gestión de proyectos de ciencia de datos

FASE 1 – Involucra la definición y planificación del proyecto, y se prioriza un enfoque tradicional, ya que se siguen los lineamientos establecidos por el PMBOK y la metodología CRISP-DM. Esta fase se divide en dos componentes. La primera, Pre-Proyecto, donde el Jefe de Proyecto debe preparar un documento formal que indique objetivos, alcance y restricciones que permita justificar o no la realización del proyecto y definir las partes interesadas, internas y externas al sector estrechamente vinculado con el proyecto, de forma de lograr comprometer, involucrar y recibir apoyo de todos los interesados. En esta etapa se vincula estrechamente la Gestión de Alcance del PMBOK y Comprensión del Negocio de CRISP-DM. En la segunda componente, Visión/Definición, el Jefe de Proyecto debe determinar planes dentro del proyecto y para cada uno: aprobar y certificar el grado de compromiso de todos los interesados, para con el proyecto; comunicar los planes a todos los interesados; y generar espacio para el intercambio de experiencias. En esta segunda componente se involucra la gestión del tiempo, costos, calidad, riesgos, del PMBOK y la Comprensión del Negocio y la Comprensión de los datos de CRISP-DM.

FASE 2 – Involucra la ejecución y control del proyecto, y se prioriza un enfoque ágil porque tienen predominio la dinámica y adaptabilidad, características particulares de los proyectos de ciencia de datos. La presente fase presenta sus componentes solapados entre la metodología SCRUM y las fases de Preparación de los datos, Modelado, Evaluación e Implantación de la metodología CRISP-DM. Esto se debe a que al aplicar SCRUM como principal metodología ágil de gestión, los componentes Preparación y Desarrollo Colaborativo están fuertemente enlazados. En esta fase, el Jefe de Proyecto debe vigilar parte de los requisitos que se

corresponden con la formación y crecimiento, tales como: grado de cooperación y comunicación, crecimiento del equipo y desarrollo de habilidades; y la mayoría de los que se corresponden a los procesos internos. Esta fase se divide en tres componentes. La primera, Preparación, donde el Jefe de Proyecto supervisa las etapas de Preparación de los Datos de CRISP-DM y luego definir los planes graduales de lanzamientos correspondientes a cada requerimiento que tendrá como resultado un modelo entregable. La segunda es Desarrollo Colaborativo; la razón del nombre es que la palabra colaboración implica trabajo en equipo, comunicación constante y permanente participación de las partes interesadas, principalmente del o los usuarios representativos. Esto es muy importante para garantizar tanto la culminación de cada modelo entregable como del proyecto en su conjunto. Al cumplir el tiempo asignado para terminar el entregable, se procede a evaluar y probar el mismo (Fase de Evaluación de CRISP-DM). Mientras se ejecuta al *Plan de Lanzamiento*, se efectúan *Revisiones* que tienen como objetivo demostrar concretamente y claramente el progreso del equipo y, recibir retroalimentación de los usuarios sobre el entregable generado (SCRUM).

Al finalizar cada *Plan de Lanzamiento* se genera una reunión de *Retrospectiva* que tiene como objetivos: escuchar distintos puntos de vista dentro del Equipo de Proyecto, identificar colaborativamente las causas de los principales problemas del equipo durante la ejecución del *Plan de Lanzamiento* e idear, consensuar y seleccionar acciones de mejoras concretas que pueda ejecutar el equipo en el próximo *Plan de Lanzamiento* (SCRUM). Recopilar lecciones aprendidas a partir de las reuniones diarias (Ceremonias de SCRUM) donde el *Equipo de Proyecto* le responde al *Director de*

Proyecto las tres preguntas clásicas utilizadas en SCRUM: *¿Cuál fue mi avance desde la última reunión diaria?*, *¿En cuáles tareas me comprometo a trabajar hasta la próxima reunión diaria?* y, *¿Qué problemas tengo que me frenan o bloquean?* La tercera componente, Adaptación, Dirección, Monitoreo y Control, debe integrar las componentes, visión/definición, preparación y desarrollo colaborativo. El objetivo de esta componente es garantizar que no se establecerán cambios incontrolados, es decir todos los cambios que se produzcan deben tener como objetivo mantener la integridad del proyecto a lo largo de todo su desarrollo.

FASE 3 – involucra la evaluación final y cierre, y si bien, no hay un preeminencia de un enfoque sobre otro, ya que ambos, tradicional y ágil tienen como objetivos formular un final claro del proyecto, se plantea el cierre teniendo en cuenta los lineamientos del PMBOK. Esta fase se divide en dos componentes: El cierre, cuyo objetivo es formular un final claro del proyecto, firmando la aceptación del mismo y proporcionando toda la documentación necesaria.

La segunda componente, Pos-Proyecto, Mantenimiento y Mejora Continua, tiene como objetivo asegurar que la solución sigue aportando valor para la organización.

Para el caso de una mejora significativa o una actualización, la propuesta metodológica de gestión de proyectos con enfoque “híbrido”, vuelve a iniciarse a la Fase 1: Definición y Planificación.

La propuesta obedece al requisito de mantener la documentación justa y definir técnicas de difusión de políticas de documentación, el nombre del proyecto, donde se guardan todos los documentos vinculados al proyecto y tienen acceso al mismo todas las partes interesadas. De esta forma, se logra una amplia difusión de toda la documentación.

5. Validación de la propuesta a un caso de estudio real

Con el fin de validar la propuesta metodológica se empleó en la gestión de un proyecto perteneciente al GIBD (Grupo de Investigación de Bases de Datos) de la Facultad Regional Concepción del Uruguay de la Universidad Tecnológica Nacional. Dicho proyecto se encuentra en el marco de una línea de investigación sobre Agenda Setting, concepto que hace referencia a la influencia que tienen los medios de comunicación en la fijación de temas en la opinión pública. En dicha línea de investigación se busca realizar un análisis de dominio cruzado para medir los efectos de la instalación de asuntos en la agenda pública tomando como base artículos escritos sobre diferentes temáticas en medios digitales de relevancia para determinar los

tópicos que tratan y luego analizar su difusión en redes sociales empleando técnicas de minería de textos y procesamiento de lenguaje natural. Como red social a estudiar se eligió Twitter, por lo que es necesario recolectar y preprocesar tweets en función de distintos parámetros en relación a las noticias de los diarios. En el marco de dicha investigación, se planteó como subproyecto el *diseño y desarrollo de una aplicación web para la recolección y preprocesamiento de tweets de forma automática*, según determinados parámetros que funcionan como filtros, que pueden aplicarse al momento de la captura o durante un proceso posterior.

La arquitectura completa propuesta (Figura 4) sigue el modelo cliente-servidor: del lado del cliente se utiliza el framework de JavaScript Vue.js en conjunto con el framework de estilos por componentes Vuetify. Este cliente se comunica con un servidor compuesto por una API desarrollada en Python, utilizando el micro-framework Flask, que funciona como interlocutor de los diferentes servicios utilizados. Para la persistencia de los datos de procesos se utiliza MongoDB, para la persistencia de los tweets capturados, Amazon Web Service 3, y el gestor de tareas Celery, para ejecutar las tareas de recolección, procesado y almacenamiento en segundo plano. En la comunicación con Celery se utiliza el motor de base de datos Redis para mantener una cola de mensajes, con el objetivo de iniciar o detener las tareas. Estos servicios se despliegan en containers Docker, junto con interfaces de usuario, para su mejor monitoreo y control.

Para el desarrollo del subproyecto mencionado (desarrollo de aplicación para captura y preprocesamiento de tweets) se empleó la propuesta metodológica de enfoque “híbrido”. A continuación se describe la implementación de las distintas fases:

FASE 1 – Definición y planificación

Pre-Proyecto

Luego de las reuniones informales en las cuales se acordó el desarrollo del proyecto, se formalizó el mismo en un documento que indica objetivos, alcance y restricciones. Dicho documento justifica la realización del proyecto y define las partes interesadas. En la Tabla 4 se presenta un resumen del contenido de dicho documento.

Tabla 4. Documento formal de la etapa de pre-proyecto

Dimensión	Descripción
Objetivo	Diseño e implementación de una aplicación web para la recolección de tweets de forma automática según determinados parámetros
Alcance	El desarrollo del proyecto implica el diseño de la solución, la investigación y elección de las distintas herramientas con su posterior inducción, la configuración de las mismas, la preparación de los entornos y el desarrollo de la aplicación web y la API que le brinde servicio
Restricciones	Uso de herramientas open source. Para la persistencia de los tweets capturados se utilizará un servicio de almacenamiento en la nube, hasta contar con la infraestructura requerida para poder realizarlo localmente.

Visión/Definición

Se realizó un estudio del negocio y de los datos, lo que permitió comprender con mayor detalle el objetivo del proyecto, puntualmente qué datos se necesitan capturar y el preprocesamiento que se le debe aplicar a los mismos.

Adicionalmente se determinó que la aplicación realizará la recolección de tweets en base a filtros predefinidos por el usuario antes de iniciar la tarea. En particular, se especifican palabras clave o hashtags que deben aparecer en el tweet para que se capture y almacene. Mediante el uso de la Search API de Twitter, se recolectan tweets a partir de los filtros dados por el usuario. Dicha petición devuelve un archivos JSON con 75 variables de cada tweet, y luego de un primer preprocesamiento, se almacenan en MongoDB las siguientes variables: *created_at*, *timestamp_ms*, *favorite_count*, *in_reply_to_user_id*, *coordinates*, *geo*, *source*, *lang*, *variables del usuario*, *hashtags*, *symbols*, *user_mentions*, *urls*, *contributors*, *text*, *quote_count*, *retweeted*, *reply_count*, *retweet_count*, *id*, *favorited*.

Se determinó el plan del proyecto teniendo en cuenta la metodología a seguir. El diseño del plan del proyecto involucró la planificación de tiempo, costos, recursos humanos y riesgos. Se realizó la comunicación y aprobación del plan con todos los interesados.

También se definieron las herramientas a utilizar y la interacción entre las mismas, haciendo énfasis en la escalabilidad y eficiencia del sistema, que se han mencionado al inicio de esta sección.

FASE 2 – Ejecución y control

Preparación

El Jefe de Proyecto definió en conjunto con el equipo de trabajo los planes graduales de lanzamientos correspondientes a cada requerimiento, que tienen como resultado un modelo entregable (Tabla 5).

Desarrollo Colaborativo

Se realizó el desarrollo de los planes graduales, trabajando en equipo, con comunicación constante y permanente participación de las partes interesadas, principalmente del desarrollador y de los usuarios representativos.

Se mantuvieron reuniones diarias para analizar el avance logrado, las próximas tareas y los problemas actuales. Estas reuniones permitieron recopilar lecciones aprendidas.

Al llegar a cada entregable, se fueron realizando las pruebas y evaluaciones de los mismos, junto con revisiones del trabajo realizado para demostrar el progreso del equipo y recibir retroalimentación de los usuarios sobre el entregable generado.

Tabla 5. Etapas del proyecto y entregables

Etapa	Entregable
Diseño de propuesta de solución.	Esquema de solución.
Investigación y selección de los lenguajes y herramientas.	Documento que detalla las características de los lenguajes y herramientas a utilizar.
Diseño de la arquitectura de solución	Arquitectura de solución con una descripción de la interacción entre las herramientas.
Diseño y construcción de entornos de desarrollo y producción reproducibles	Entornos de desarrollo y producción
Instalación de las bases de datos	Bases de datos instaladas y configuradas
Desarrollo de la API	API web
Desarrollo de la aplicación web	Aplicación web

Adaptación, Dirección, Monitoreo y Control

Como se mencionó anteriormente, se realizó un proceso de dirección y control constante. Esto permitió realizar las adaptaciones necesarias durante el transcurso del proyecto de forma controlada.

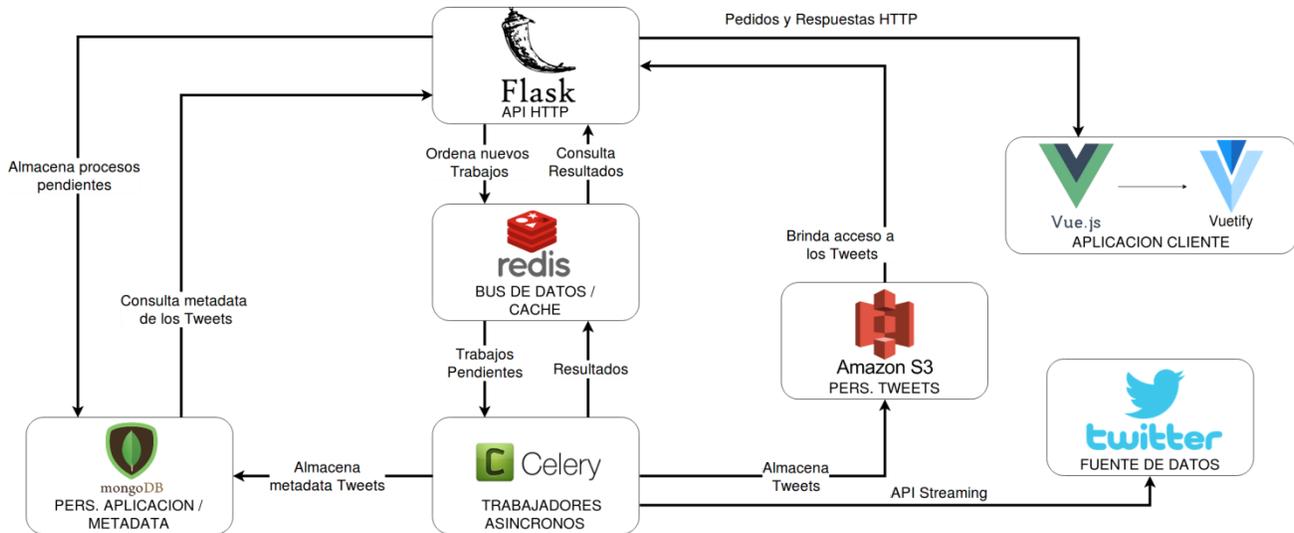


Figura 4. Arquitectura de la solución completa propuesta

FASE 3 – Evaluación final y cierre

Cierre

Se realizó el cierre formal de la primera etapa del proyecto. El proyecto continúa para realizar mejoras a la aplicación desarrollada.

Pos-Proyecto, Mantenimiento y Mejora Continua

Actividades en desarrollo para asegurar que la solución siga aportando valor a la organización. En esta misma línea, se pactó volver a la Fase 1 para plantear una actualización de la aplicación.

Documentación, organización y coordinación del equipo

Se creó un repositorio para almacenar el código y la documentación del proyecto, dando acceso al mismo todas las partes interesadas. La organización del equipo y las actividades se realizó por medio de Trello² y Slack³. Esto permitió tener los objetivos claros, una correcta gestión del tiempo y una comunicación fluida entre los miembros del equipo.

6. Resultados

La Propuesta Metodológica de enfoque “híbrido”, permite mantener la dinámica entre equipo de proyecto y los interesados para con las actividades afectadas a cada uno. Por otro lado se logró un buen equilibrio entre la propuesta establecida en el trabajo anterior a partir de la obtención de los puntos fuertes de la guía de buenas

prácticas, PMBOK, ATERN y la metodología ágil ampliamente utilizada en proyectos vinculados con las tecnologías de la información, SCRUM con la metodología tradicional CRISP-DM, específicamente desarrollada para gestionar proyectos de ciencia de datos.

7. Conclusiones y Trabajo Futuros

Los resultados obtenidos permiten afirmar que la propuesta metodológica de enfoque “híbrido” ha cumplido con los objetivos propuestos. En este trabajo se presenta la validación de la propuesta para un proyecto de ciencia de datos en el que se construye una aplicación que permite la recolección y preprocesamiento de datos de fuentes externas (twits) para la confección de la base de datos a analizar. Los aportes más relevantes de la aplicación de esta propuesta metodológica a este proyecto particular fueron la agilización en la gestión y desarrollo del proyecto. Esto se logró mediante de la colaboración entre las partes interesadas, logrando integrar áreas vinculadas con el proyecto, a través del trabajo en conjunto de interesados y equipo del proyecto. Esta propuesta de enfoque “híbrido” que integra las fortalezas de metodologías tradicionales y ágiles, resulta atractiva para proyectos de minería de datos, dado que: permite optimizar el tiempo y recursos utilizados, no requiere formación específica por su sencillez, sólo se genera la documentación necesaria y la información requerida para asegurar una comunicación eficiente y no incurrir en los mismos errores a partir de las lecciones aprendidas. Luego de la ejecución de este proyecto empleando la metodología de enfoque “híbrido” se pudo concluir que la misma resulta aplicable a proyectos de ciencia de datos en general, ya que no sólo se orienta a proyectos que

² <https://trello.com/>

³ <https://slack.com/>

construyen modelos o patrones estáticos a partir del análisis de datos del negocio, sino que considera fuertemente escenarios en los que la base de datos a analizar se actualiza continuamente, mediante la incorporación de datos de fuentes externas, y el dinamismo de los modelos es clave para la toma de decisiones en tiempo real.

Se considera como trabajo futuro ampliar la validación de la propuesta metodológica a nuevos proyectos de ciencia de datos, para realizar validaciones en distintos escenarios.

8. Referencias

- [1] Ngai, E. W.; Xiu, L.; Chau, D. C. "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert systems with applications*, vol. 36, no. 2, pp. 2592-2602, (2009).
- [2] Witten, I. H.; Frank, E., Hall, M. A.; Pal, C. J. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, (2016).
- [3] Chapman, P., "CRISP-DM 1.0 Step-by-step data mining guide," (2000).
- [4] Mariscal, G.; Marban, O.; Fernandez, C. "A survey of data mining and knowledge discovery process models and methodologies," *The Knowledge Engineering Review*, vol. 25, no. 2, pp. 137-166, (2010).
- [5] Varajão, J. Success Management as a PM knowledge area – work-in-progress. *Procedia Computer Science*. Vol. 100, pp.1095 – 1102, (2016).
- [6] Colomo-Palacios, R.; Casado-Lumbreras, C.; Soto-Acosta, P.; García-Peñalvo, F.J.; Tovar-Caro, E. Project managers in global software development teams: a study of the effects on productivity and performance, *Software Quality*. 22 (1)3–19, (2014).
- [7] Varajão, J. Dominguez, C. Ribeiro, P. Paiva, A. Critical success aspects in Project management: similarities and differences between the construction and software industry, *Tech. Gazette* 21 (3) 583–589, (2014).
- [8] Montequin, S.; Fernandez, C.; Fernandez, O.; Balsera, J.V. [Analysis of the Success Factors and Failure Causes in Projects: Comparison of the Spanish Information y Communication Technology \(ICT\) Sector](#). *Journal Information Technology Project Management*, Vol. 7(1), pp. 18-31, (2016).
- [9] Chow, T.; Chao, D. A survey of critical success factors in agile software projects. *Journal of systems and software*, Jun. Available: Science Direct, 81(6):961–971, (2008).
- [10] Lehtinen, T.; Mäntylä, M.; Vanhanen, J.; Itkonen, J.; Lassenius, C. Perceived causes of software project failures – An analysis of their relationships. *Information and Software Technology* 56, pp. 623–643, (2014).
- [11] Ramos, P.; Mota, C. Perceptions of success and failure factors in information technology projects: a study from Brazilian companies. *Procedia - Social and Behavioral Sciences* 119, pp. 349 – 357, (2014).
- [12] Elkadi, H. Success and failure factors for e-government projects: A case from Egypt. *Egyptian Informatics Journal*, 14, pp. 165–173, (2013).
- [13] ElEmam, K.; Koru, A.G. A replicated survey of IT software project failures *Software*. *IEEE Software*, vol. 25, pp. 84–90, (2008).
- [14] Bentley, L.; Whitten, J. *Systems Analysis and design methods*. MacGraw Hill, pp. 9–11, (2007).
- [15] Cristaldo, P.; Ballejos, L.; Alé, M. Un enfoque híbrido de gestión de proyectos de TICs en el sector público. 44 *Jornadas Argentinas de Informática*. Universidad Nacional de Rosario, Facultad de Ingeniería, Argentina, ISBN, (2015).
- [16] Spundak, M. Mixed agile/traditional project management methodology – reality or illusion? *Procedia - Social and Behavioral Sciences* 119, pp. 939 – 948, (2014).
- [17] Conforto, E. C.; Amaral, D. C. Evaluating an agile method for planning and controlling innovative projects. *Project Management Journal*, 33(4), pp. 4–14, (2008).
- [18] Schwaber, K.; Beedle, M. *Agile Software Development with SCRUM*. In: Conchango ISBN: 0130676349, (2006).
- [19] Schwaber, K.; Sutherland, J. *The SCRUM Guide, the Definitive Guide to SCRUM: The Rules of the Game*. http://www.SCRUM.org/Portals/0/Documents/SCRUMguides/SCRUM_Guide.pdf, (acceso 14.02.17) (2011).
- [20] Böhm, A. *Application of PRINCE2 and the Impact on Project Management*. 1 Ed. Grin Publishing, 60 p. ISBN-10: 3640424247, ISBN-13: 978-3640424245, (2009).
- [21] *PMBOK Guide*. Project Management Institute. A Guide to the Project Management Body of Knowledge. 5 Ed. NewtownSquare, Pennsylvania 19073-3299 USA. PE: Project Management Institute Inc .ISBN: 978-1-935589-67-9, (2013).
- [22] Nogueira, D. *Agile Data Mining: Uma metodologia ágil para o desenvolvimento de projetos de data mining*. Faculdade de Engenharia da Universidade do Porto, (2014).
- [23] Alnoukari, M.; Alzoabi, Z. Hanna, A. Applying Adaptive Software Development (ASD) Agile Modeling on Predictive Data mining Applications: ASD-DM Methodology. *Information Technology, ITSIM 2008*. International Symposium on (Volume 2). doi: 10.1109/ITSIM.2008.4631695, (2008).

- [24] Santana Do Nascimento, G.; Aparecida de Oliveira, A. Agile KDD An Agile Knowledge Discovery in Databases Process Model. Federal University of Sergipe São Cristóvão, Brazil gsnascimento@petrobras.com.br Adicinéia Federal University of Sergipe São Cristóvão, Brazil adicineia@ufs.br (2013).
- [25] Guo, P. Data Science Workflow: Overview and Challenges. Acceso acm.org. (2013).
- [26] Ibm. Have you seen ASUM-DM? <https://developer.ibm.com/predictiveanalytics/2015/10/16/have-you-seen-asum-dm/>. (2015).
- [27] Way, R. Model deployment: the moment of truth, Corios Redpaper, (2013).
- [28] Niaksu, O. Development and application of Data Mining methods in medical diagnostics and healthcare management. Vilnius University, (2015).
- [29] Mariscal, G.; Marbán, O.; Segovia, J. Un enfoque Ágil para el Desarrollo de Proyectos de Data Minig, Researchgate, Marchx, G. (2014). Big Data and Agile: The Perfect Marriage. Head of Agile Competency Center, EPAM Canadá. (2013).
- [30] Marchx, G. Big Data and Agile: The Perfect Marriage. Head of Agile Competency Center, EPAM Canadá, (2014).
- [31] Lopuszynski, M. Agile Approach to Data Minig Projects. Warsaw Data Science Meetup, (2016).
- [32] Marinescu, M., Traianpele, D. Modelling the strategic success factors of the Romanian ICT based Companies. Procedia - Social and Behavioral Sciences 58, pp.1111 – 1120, (2012).
- [33] Rivard, S.; Dupre, R. Information systems project management in PMJ: a brief history, Project Management Journal 40, pp. 20–30, (2009).
- [34] Lock, D. Project management. 9 Ed. Aldershot: Gower. 520 p. ISBN-13: 978-0-566-08772-1 (pbk) 978-0-566-08769-1 (hardback), (2007).
- [35] Batra, D.; Xia, W., VanderMeer, D., y Dutta, K. Balancing Agile and Structured Development Approaches to Successfully Manage Large Distributed Software Projects: A Case Study from the Cruise Line Industry, Communications of the Association for Information Systems, Vol. 27, Article 21, p. 379-394, (2010).
- [36] Stare, A. Agile Project Management in Product Development Projects. Procedia - Social and Behavioral Sciences 119, 295 – 304, (2014).
- [37] Fernandez, D. y Fernandez, J. Agile Project Management: Agilism Versus Traditional Approaches. Journal of Computer Information Systems, Vol. 49, Nro. 2, p. 10-17. (2008).
- [38] Karlström, D.; Runeson, P. Combining Agile Methods with State-Gate Project Management. IEEE Software, Vol. 22, Nro. 3, p. 43-49. (2005).
- [39] Saltz, J. The Need for New Processes, Methodologies and Tools to Support Big Data Teams. International Conference on Big Data, IEEE, Santa Clara, CA, USA (2015).
- [40] KDnuggets, “Crisp-dm, still the top methodology for analytics, data mining, or data science projects” <http://www.kdnuggets.com/2014/10/crisp-dm-top-ethodology-analytics-data-mining-data-science-projects.html>, (2014)
- [41] Highsmith, J. Agile project management: creating innovative products. 2nd ed. Boston, MA: Addison-Wesley. 432 p. (2010).
- [42] Charvat, J. Project management methodologies: selecting, implementing and supporting methodologies and processes for projects. Hoboken, NJ: Wiley. 264 p. (2003).