

A Process Model for Data Mining Projects

Un Modelo de Procesos para Proyectos de Explotación de Información

Juan A. Vanrell¹, Rodolfo Bertone², Ramón García-Martínez³

¹ Escuela de Postgrado, Universidad Tecnológica Nacional. Ciudad Autónoma de Buenos Aires, Argentina. javanrell@gmail.com.

² Facultad de Informática. Universidad Nacional de La Plata. La Plata, Argentina. pbertone@lidi.info.unlp.edu.ar

³ Grupo de Investigación en Sistemas de Información. Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús. Remedios de Escalada, Argentina. rgarcia@unla.edu.ar.

INFORMACIÓN DEL ARTÍCULO

Tipo de artículo: Investigación

Historia del artículo:

Recibido: 23/04/2012

Correcciones: 30/05/2012

Aceptado: 31/05/2012

Palabras clave

Explotación de información, modelo de procesos.

Categories and Subject Descriptors:

D.2.1 [Requirements/Specifications]:
Methodologies.

General Terms

Documentation

Keywords

Data mining projects, process model.

ABSTRACT

Data mining projects have very different characteristics from traditional software development projects. Classics steps of analysis, design, development, integration and testing do not fit the natural steps of development process of such projects. In this context, we developed a process model for data mining projects for SMEs following the guidelines of the process model for the software industry (COMPETISOFT).

RESUMEN

Los proyectos de explotación de información poseen características muy distintas a las de los proyectos de desarrollo de software tradicionales. Las clásicas etapas de análisis, diseño, desarrollo, integración y testeo no encajan con las etapas naturales de los procesos de desarrollo de este tipo de proyectos. En este contexto, se desarrolló un modelo de procesos para proyectos de explotación de información para PyMEs siguiendo los lineamientos del modelo de procesos para la industria de software (COMPETISOFT).

1. INTRODUCCIÓN

Actualmente existen en el mercado distintos modelos que ayudan a llevar a cabo proyectos con un nivel de calidad esperado en forma repetitiva como pueden ser el de la norma ISO/IEC15504 e ISO/IEC12207, el modelo CMM y su versión actual CMMI [1], MoProSoft [2] o su versión iberoamericana COMPETISOFT [3]. Todos estos son modelos genéricos por lo cual pueden ser utilizados para la ejecución de cualquier tipo de proyecto [4-5].

Dentro de los distintos proyectos que son llevados a cabo por empresas dedicadas al área de tecnologías de la información se encuentra un conjunto denominado proyectos de explotación de información. Como todo conjunto posee características propias que lo hacen diferenciarse del resto. Se considera que estas características son lo suficientemente significativas como para justificar la construcción de un modelo de procesos que se ajuste a este tipo de proyectos [6].

Siguiendo los lineamientos de los creadores de MoProSoft y COMPETISOFT para la definición de un modelo que pueda ser utilizado por pequeñas y medianas empresas (Pymes) que serán: fácil de entender, fácil de aplicar y relativamente económicos en su implementación, se decidió crear un modelo de procesos de explotación de información orientado a Pymes tomando como base el modelo COMPETISOFT y adecuándolo a los procesos utilizados para los proyectos de explotación de información.

En este contexto haremos a continuación una introducción a COMPETISOFT y CRISP-DM (sección 2), para luego presentar los problemas encontrados (sección 3), la solución propuesta (sección 4) y un caso de estudio (sección 5) incluyendo las principales conclusiones sobre este trabajo (sección 6).

2. COMPETISOFT Y CRISP-DM

En esta sección se introduce el modelo de procesos COMPETISOFT (sección 2.1) y la metodología de explotación de información CRISP (sección 2.2).

2.1 Competisoft

COMPETISOFT es la proyección a nivel iberoamericano del modelo de procesos para el desarrollo de software MoProSoft creado por encargo de la Secretaría de Economía Mexicana para servir de base a la norma Mexicana para la Industria de Desarrollo y Mantenimiento de Software. El modelo inicial fue modificado y adecuado a las necesidades de otros países, se le incorporó el modelo de evaluación EvalProSoft [7] y se definieron niveles de madurez.

Su propósito es fomentar la estandarización de las operaciones de pequeñas y medianas empresas o departamentos internos de desarrollo, a través de la incorporación de las mejores prácticas en gestión e ingeniería de software, esperando "elevar la capacidad de las organizaciones para ofrecer servicios con calidad y alcanzar niveles internacionales de competitividad". El modelo busca ser fácil de entender, fácil de aprender, no

costoso en su adopción y ser la base para alcanzar evaluaciones exitosas con otros modelos o normas como ISO/IEC 15504, ISO/IEC 12207 o CMM.

Este modelo puede ser utilizado tanto por organizaciones que no cuenten con procesos establecidos, ajustándolo de acuerdo a sus necesidades, como por organizaciones que ya poseen procesos establecidos que pueden utilizar como punto de referencia para identificar los elementos faltantes.

Además de definir procesos COMPETISOFT define un patrón que debe ser utilizado para documentar aquellos procesos que una empresa requiere agregar a los existentes en el modelo o para documentar la adecuación de los que ya se encuentra en el mismo. Dicho patrón se encuentra constituido por tres partes: Definición general del proceso, Prácticas y Guías de ajuste. El modelo a desarrollar pretende seguir este patrón para la documentación de los procesos de explotación de información.

La estructura del modelo se encuentra dividida en tres categorías: Alta Dirección (DIR), Gerencia (GER) y Operaciones (OPE) reflejando la estructura de una organización. Estas categorías contienen los procesos de gestión de negocio (DIR), gestión de procesos, gestión de proyectos y gestión de recursos (GER) y administración de un proyecto específico, desarrollo de software y mantenimiento de software (OPE).

En palabras de los creadores de COMPETISOFT la categoría de Alta Dirección es la "categoría de procesos que aborda las prácticas de Alta Dirección relacionadas con la gestión del negocio" y "proporciona los lineamientos a los procesos de la Categoría de Gerencia y se retroalimenta con la información generada por ellos", la categoría de gerencia es la "categoría de procesos que aborda las prácticas de gestión de procesos, proyectos y recursos en función de los lineamientos establecidos en la Categoría de Alta Dirección", además "proporciona los elementos para el funcionamiento de los procesos de la Categoría de Operación, recibe y evalúa la información generada por éstos y comunica los resultados a la Categoría de Alta Dirección" y la Categoría de Operación es la "categoría de procesos que aborda las prácticas de los proyectos de desarrollo y mantenimiento de software", además "esta categoría realiza las actividades de acuerdo a los elementos proporcionados por la Categoría de Gerencia y entrega a ésta la información y productos generados".

2.2 CRISP-DM

La metodología CRISP-DM [8] se encuentra definida en base a un modelo jerárquico de procesos. El foco se pondrá en los procesos del nivel superior que son lo suficientemente genéricos como para cubrir todas las posibles aplicaciones de explotación de información.

Esta metodología define un ciclo de vida de los proyectos de explotación de información que define las principales fases de un proyecto de este tipo. Estas fases

son: Entendimiento de Negocios, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue. Claramente estas fases difieren de las fases definidas para un proyecto de desarrollo de software clásico (inicio, requerimientos, análisis y diseño, construcción, integración y pruebas y cierre). A continuación se presenta el concepto de cada una de las fases identificadas por CRISP-DM.

En la fase de Entendimiento del Negocio se deben entender los objetivos del proyecto y los requerimientos desde una perspectiva del negocio y luego convertir este conocimiento en una definición de un problema de explotación de información y diseñar un plan preliminar para lograr dichos objetivos.

El Entendimiento de los Datos comienza con la recolección inicial de datos y procede con las acciones para familiarizarse con ellos, identificar problemas de calidad, identificar primeras pautas en los datos o detectar subconjuntos interesantes de las hipótesis de información oculta.

La fase de Preparación de los Datos cubre todas las actividades para construir el conjunto de datos final desde los datos iniciales, las tareas de esta fase pueden ser realizadas muchas veces y sin un orden preestablecido, incluye tanto la selección de tablas, registros y atributos como transformación y limpieza de datos para herramientas de modelado.

El Modelado incluye la selección de técnicas de modelado y la calibración de sus parámetros a los valores óptimos, suelen existir distintas técnicas para un mismo problema de explotación de información y cada una de ellas suele tener ciertos requisitos sobre los datos, muchas veces es necesario volver a la fase de preparación de los datos.

La Evaluación requiere la construcción de uno o varios modelos que aparentan tener la mayor calidad desde una perspectiva de análisis, requiere la evaluación del modelo y revisión de los pasos ejecutados para la construcción del modelo para asegurarnos de lograr los objetivos de negocio, al final de esta fase se debería poder tomar una decisión respecto de la utilización de los resultados.

Por último, la fase de despliegue puede ser tan simple como generar un reporte o tan compleja como implementar un proceso de explotación de información repetible a través de toda la empresa.

3. PROBLEMA

Al desarrollar diferentes proyectos de explotación de la información con un alto grado de previsibilidad y calidad se utilizan distintos modelos de producción y metodologías. Estas herramientas permiten controlar la calidad final de producto a desarrollar estableciendo controles sobre cada una de las etapas que interviene el en proceso productivo, entendiendo por proceso productivo no solo la producción en sí misma, sino

también las tareas relacionadas a la gestión de un proyecto y de la empresa que lo desarrolla. En el caso de proyectos clásicos existen, como se mencionó al principio de esta Tesis, modelos bien probados como puede ser CMM o el modelo para PyMEs COMPETISOFT. Estos modelos han sido utilizados en suficientes proyectos de forma que pueden ser considerados modelos estables y altamente testeados (en el caso de COMPETISOFT se puede considerar al modelo MoProSoft como el más testado). Sin embargo, se considera que estos modelos no son adecuados para empresas que se dedican a llevar a cabo proyectos de explotación de información.

Por otro lado existen metodologías que acompañan el desarrollo de proyectos de explotación de información entre las cuales destacamos a CRISP-DM, P³TQ y SEMMA que si bien fueron probadas y tienen un buen nivel de madurez en cuanto al desarrollo del proyecto, dejan de lado aspectos a nivel gestión de proyectos y de empresa.

Aceptando que los proyectos de desarrollo de software tradicional poseen características muy distintas a los proyectos de explotación de información, sobre todo en la parte operativa de un proyecto [9], una lectura rápida a la documentación del modelo de procesos de desarrollo de software, COMPETISOFT, muestra grandes diferencias en cuanto a los procesos naturales de los proyectos de explotación de información. La diferencia más significativa se presenta en los procesos de desarrollo y mantenimiento de software en los cuales COMPETISOFT define como proceso natural el ciclo de fases de un proyecto de software tradicional. Las fases de Inicio, Requisitos, Análisis y Diseño, Construcción, Integración, Pruebas y Cierre no resultan naturales en un proyecto de explotación de información.

En la misma línea, al evaluar las principales metodologías existentes para los proyectos de explotación de información mencionados anteriormente, se observa la falta de herramientas que permitan soportar de forma completa la fase de administración de proyectos. Esta fase se encuentra bien definida y agrupada en el proceso de "Administración de Proyectos Específicos" dentro de la metodología COMPETISOFT. Sin embargo se encuentran algunas definiciones de tareas o actividades que pertenecen a este proceso como puede ser la construcción de un plan de desarrollo, definir el protocolo de entrega con el cliente, establecer el equipo de trabajo, definir el plan de manejo de riesgos,...

Teniendo en cuenta estos dos problemas esta Tesis se orienta a acercar los procesos de la categoría de operación definidos en COMPETISOFT con los procesos requeridos por los proyectos de explotación ya sea adecuando las fases definidas en COMPETISOFT a las necesarias para los proyectos de explotación de información, eliminando las fases que no son adecuadas o proponiendo nuevas fases en su reemplazo. En este trabajo se utiliza como metodología de referencia a CRISP-DM dado que, al comparar las tres metodologías

mencionadas anteriormente [10], se pudo observar que esta última contenía todos los elementos a nivel operación de las otras.

4. SOLUCIÓN PROPUESTA

Se propone como solución a los problemas mencionados en la sección anterior, mantener los procesos de la categoría de Operación definidos en COMPETISOFT, como Administración de Proyectos y Desarrollo de Proyectos readecuándolos a los proyectos de explotación de información.

Las categorías de Alta Dirección y Gerencia se mantienen invariables basándose en el modelo COMPETISOFT dado que se definen a un nivel empresarial y no a un nivel de producción con lo cual pueden ser adecuadas para cualquier tipo de proyecto. Como se menciona anteriormente la metodología CRISP-DM no hace distinción entre procesos sino que posee un único proceso en el cual se definen todas las tareas dividiéndolas en fases, ya sea las relacionadas con la administración como las relacionadas con el desarrollo.

Esta división es necesaria ya que ambos procesos deben ser ejecutados en forma concurrente, por un lado se debe desarrollar el proyecto propiamente dicho y al mismo tiempo se deben ejecutar tareas de administración del proyecto para controlar su avance, realizar correcciones y recolectar datos para futuros proyectos.

La reestructuración de estos dos procesos en el desarrollo de proyectos de Explotación de Información brindará mayor claridad a las tareas de administración ya que, como se menciona anteriormente, las tareas de administración que se encuentran mencionadas en las metodologías evaluadas se encuentran dentro de un mismo proceso de desarrollo.

Esta división no se limita a separar las tareas existentes en las metodologías sino que se construye el nuevo proceso basándose en el proceso de administración utilizado para el desarrollo de software clásico, específicamente el definido en el modelo COMPETISOFT, lo cual enriquece el proceso de administración de proyectos. A este proceso se lo adecuó para responder a las exigencias de los proyectos de Explotación de Información.

La incorporación del proceso de administración del proyecto, y su concepción como un proceso distinguible del de desarrollo, constituye una aportación a la gestión de los proyectos de Explotación de Información, ya que ninguna de las metodologías evaluadas, incluyendo a CRISP-DM, hace una separación clara entre las tareas que se deben llevar a cabo para la administración del proyecto y las que se llevan a cabo para su desarrollo.

Partiendo del proceso definido para COMPETISOFT, se definió uno nuevo que incorpora las tareas definidas en las fases de CRISP-DM que se consideraron fuertemente ligadas a los procesos de administración, estas tareas

son las de "Determinar los objetivos del Negocio" que se encuentra relacionada con la elicitación de requerimientos y "Evaluación de la Situación", ambas ubicadas en el subproceso de "Planificación / Entendimiento del Negocio", y "Planear la entrega", la cual fue ubicada dentro del subproceso de "Cierre/Entrega". El resto de las tareas definidas en el proceso de administración pertenecen a COMPETISOFT. Junto con esta reestructuración de tareas se agregaron actividades para llevar a cabo cada una de las etapas definidas en el proceso de administración, para aquellos casos en los que no existía ya un conjunto de actividades definidas, como puede observarse en las tareas.

Se definieron al mismo tiempo técnicas que pueden ser aplicadas como recomendación para llevar a cabo las actividades que se mencionan en el proceso de administración de proyectos. Para el caso del proceso de desarrollo del proyecto, no se tomo como base el proceso de desarrollo definido en COMPETISOFT dado que, como se indicó en el capítulo anterior, las fases de

desarrollo de los proyectos de Explotación de Información no coinciden naturalmente con las fases mediante las cuales se desarrollan los proyectos de software clásicos.

Si bien se mantuvo la estructura utilizada por COMPETISOFT, tanto la división de los subprocesos como las tareas a realizar en cada uno de ellos, fueron definidas a partir de las fases de desarrollo planteadas por la metodología CRISP-DM. Se agregaron actividades faltantes para obtener un proceso coherente y detallado y se incorporaron a este proceso recomendaciones de las técnicas a utilizar para cada una de las tareas definidas en los subprocesos del proceso de desarrollo del proyecto.

El resultado de esta solución puede verse en las Tablas 1 y 2, en las que se omite, por falta de espacio, la descripción de las actividades necesarias para realizar cada tarea y las técnicas sugeridas.

Tabla 1. Proceso de Administración de Proyectos

SUBPROCESO	TAREA	SALIDA
Planificación / Entendimiento del negocio	Entendimiento del negocio	<ul style="list-style-type: none"> Conocimiento del negocio Objetivos del negocio Criterios de éxito
	Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento	<ul style="list-style-type: none"> Proceso Especifico (forma parte del Plan de Desarrollo)
	Definir el protocolo de entrega con el cliente	<ul style="list-style-type: none"> Plan de Entrega
	Definir ciclos y actividades con base en la descripción del proyecto y en el proceso específico	<ul style="list-style-type: none"> Proceso Especifico (forma parte del Plan de Desarrollo)
	Determinar tiempo estimado para cada actividad	<ul style="list-style-type: none"> Calendario de actividades (forma parte del Plan de Desarrollo) incorpora el tiempo estimado en el Plan de Proyecto
	Elaborar plan de adquisiciones y capacitación	<ul style="list-style-type: none"> Plan de Adquisiciones y Capacitación
	Establecer el equipo de trabajo	<ul style="list-style-type: none"> Equipo de trabajo (forma parte del Plan de Desarrollo)
	Establecer el calendario de actividades	<ul style="list-style-type: none"> Calendario de actividades (forma parte del Plan de Desarrollo)
	Calcular el costo estimado del proyecto	<ul style="list-style-type: none"> Costo estimado (forma parte del Plan de Proyecto)
	Evaluación de la situación	<ul style="list-style-type: none"> Inventario de recursos Requerimientos, suposiciones y restricciones Riesgos y contingencias (forma parte del Plan de Proyecto nombrado como Plan de Manejo de Riesgos) Terminología Costos y beneficios
Producir un Plan de Proyecto	<ul style="list-style-type: none"> Plan de Proyecto, incluye ciclos y actividades, tiempo estimado, plan de adquisiciones y capacitación, equipo de trabajo, costo estimado, calendario, plan de manejo de riesgos y protocolo de entrega 	
Producir un Plan de Desarrollo	<ul style="list-style-type: none"> Plan de Desarrollo (incluye descripción del producto y entregables, proceso específico, equipo de trabajo y calendario) Lista inicial de técnicas y herramientas 	
Formalizar el inicio de un nuevo ciclo del proyecto		
Realización	Acordar las tareas con el equipo de trabajo	
	Acordar la distribución de información	
	Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario	
	Revisar cumplimiento del plan de adquisiciones y capacitación	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Administrar subcontratos	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento Reporte de Mediciones y Sugerencias de Mejora
	Registrar costo real del proyecto	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Revisar el registro de rastreo basado en los productos de trabajo recolectados	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Revisar los productos terminados durante el proyecto	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Recibir y analizar las solicitudes de cambio del cliente	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento 	
Evaluación y Control	Evaluar el cumplimiento del plan de proyecto y plan de desarrollo	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Analizar y controlar los riesgos	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Generar el reporte de seguimiento del proyecto	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Cierre / Entrega	Formalizar la terminación del proyecto o ciclo	<ul style="list-style-type: none"> Documento de aceptación
	Llevar a cabo el cierre del contrato con subcontratistas	
	Generar el reporte de mediciones y sugerencias de mejora	<ul style="list-style-type: none"> Reporte de mediciones y sugerencia de mejoras - Lecciones Aprendidas
	Planear la entrega	<ul style="list-style-type: none"> Plan de entrega (forma parte del Plan de Proyecto nombrado como protocolo de entrega)

Tabla 2. Proceso de Desarrollo del Proyectos

SUBPROCESO	TAREA	SALIDA
Entendimiento del negocio	Determinar las metas del Data Mining	Metas del Data Mining
		Criterios de éxito del Data Mining
Entendimiento de los datos	Reunir los datos iniciales	Reporte de datos iniciales
	Describir los datos	Reporte de descripción de datos
	Explorar los datos	Reporte de exploración de datos
	Verificar la calidad de los datos	Reporte de calidad de los datos
Preparación de los datos	Tareas preparatorias	Datasets
	Seleccionar los datos	Descripción de los Datasets
	Limpiar los datos	Justificación de inclusión / exclusión
	Construir los datos	Reporte de limpieza de datos
		Atributos derivados
		Registros generados
		Datos combinados (combinación de tablas y agregaciones)
Modelado	Formatear los datos	Datos formateados
	Seleccionar la técnica de modelado	Técnica de modelado
		Suposiciones de modelado
	Generar el diseño de test	Diseño de test
	Construir el modelo	Establecimiento de parámetros
Evaluación	Evaluar el modelo	Modelos
		Descripción del modelo
		Evaluación del modelo
		Revisión de los parámetros establecidos
Evaluación	Evaluar resultados	Evaluación de los resultados de Data Mining respecto a los criterios de éxito
	Revisar el proceso	Modelos aprobados
	Determinar próximos pasos	Revisión del proceso
Entrega	Producir un reporte final	Lista de posibles decisiones
		Decisiones
		Reporte final
		Presentación final

5. CASO DE ESTUDIO

En su Tesis de Maestría, Flores [11] desarrolla un proyecto de explotación de información para detectar patrones en la producción de daños y/o averías en la cadena de distribución de la industria automotriz.

Para llevar a cabo su trabajo el autor optó por el uso de la metodología de desarrollo a CRISP-DM basándose en la independencia de esta metodología con respecto a las herramientas tecnológicas a utilizar en la explotación de datos. Por ser ésta de libre acceso, orientada al negocio y finalmente debido a ser la más completa de las metodologías evaluadas ya que incluye, además de los procesos de desarrollo, una fase preliminar dedicada al entendimiento del negocio que no es contemplada por el resto de las metodologías evaluadas.

Se utilizó el trabajo de Flores para realizar una comparación entre el uso de CRISP-DM y el uso del Modelo de Procesos propuesto como solución de esta Tesis, destacando similitudes, diferencias y las ventajas que provee el nuevo modelo. Se dividió esta comparación basándose en tres etapas en las que puede encontrarse un proyecto, ya que las actividades correspondientes a cada una de ellas pueden ser bien diferenciadas: Inicio, Ejecución y Fin.

Esta comparación puede visualizarse en la Tabla 3, en la cual se muestran los procesos del nuevo modelo junto con las actividades realizadas mediante el uso de CRISP-DM en la Tesis de Flores y las actividades que deberían ser realizadas en caso de utilizar el nuevo modelo.

6. CONCLUSIONES

Durante la investigación documental se encontraron diferencias importantes entre las PyMEs y las grandes empresas que justifican la creación de un modelo de

procesos diferenciado y se reconocieron las diferencias entre el modelo COMPETISOFT y las metodologías para proyectos de explotación de información.

Se detectaron carencias en lo referente a la gestión de proyectos y de la empresa en las metodologías utilizadas para proyectos de explotación de información, así como algunas herramientas básicas de gestión pero integradas en un único proceso de desarrollo que fueron consideradas insuficientes. Al mismo tiempo se consideró como no adecuado el proceso de desarrollo de software clásico para llevar a cabo proyectos de explotación de información.

Como solución a estos problemas se propuso el uso de COMPETISOFT como punto de partida para la construcción de un nuevo modelo de procesos que supla estas carencias utilizando CRISP-DM para readecuar los procesos reemplazando completamente el proceso de desarrollo definido en COMPETISOFT por el definido en CRISP-DM, incorporando actividades para lograr realizar cada una de las tareas del proceso y proponiendo herramientas para llevarlas a cabo.

En lo referente al proceso de administración se mantuvo el proceso definido por COMPETISOFT y se agregaron las tareas existentes en CRISP-DM con lo cual se logró independizar cada uno de los procesos. Finalmente se incorporaron al proceso de administración las actividades necesarias para la realización de cada tarea y se propusieron herramientas para llevarlas a cabo. En la evaluación de un caso de estudio se reconoció la falta de información histórica al ejecutar proyectos con CRISP-DM lo cual es solucionado en el modelo propuesto a través del proceso de Administración de Proyectos.

Tabla 3. Resumen de la Comparación

CATEGORÍA	PROCESO	SUB-PROCESO	CRISP-DM	MODELO DE PROCESOS PARA PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN	
Alta Dirección	Todos	Todos	No se contempla	Introduce los procesos definidos por Competisoft	
Gerencia	Todos	Todos	No se contempla	Introduce los procesos definidos por Competisoft	
Operación	Administración del Proyecto	Planificación/ Entendimiento del negocio	Descripción de los objetivos de negocio. Descripción del estado actual. Planteamiento de objetivos generales y criterios de éxito. Requerimientos, presunciones y restricciones. Plan de desarrollo. Evaluación de riesgos. Planificación de entregas. Plan de entrega.	Tareas propuestas por CRISP-DM + Definición de ciclos de desarrollo. Evaluación de tiempos (uso de PERT). Desarrollo del calendario de actividades. Plan de adquisición y capacitación de personal (uso de Diagramas de Gantt). Estimación de costos (DM-COMO o Técnicas empíricas). Identificación de beneficios. Planificación de planes de riesgos y contingencias. Planificación de entregas. Plan de Proyecto. Formalización del inicio de cada ciclo	
		Realización	No se contempla	Acordar tareas con el equipo de trabajo. Acordar la distribución de información. Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario. Revisar cumplimiento del plan de adquisiciones y capacitación. Administrar subcontratos. Recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo. Registrar costo real del proyecto. Revisar el registro de rastreo basado en los productos de trabajo recolectados. Revisar los productos terminados durante el proyecto. Recibir y analizar las solicitudes de cambio del cliente. Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos.	
		Evaluación y control	No se contempla	Evaluar el cumplimiento del plan de proyecto y plan de desarrollo. Analizar y controlar los riesgos. Generar el reporte de seguimiento del proyecto.	
Categoría	Proceso	Subproceso	CRISP-DM	Modelo de Procesos para Proyectos de Explotación de Información	
	Administración de Proyectos (cont.)	Cierre / Entrega	Planear la entrega. Generar el reporte de mediciones y sugerencias de mejora.	Tareas propuestas por CRISP-DM + Formalizar la terminación del proyecto o ciclo. Llevar a cabo el cierre del contrato con subcontratistas	
	Desarrollo del Proyecto / Fases Incluidas en CRISP-DM	Entendimiento del negocio	Uso del subproceso definido en CRISP-DM (incluye tareas propias de la administración mencionadas en el proceso correspondiente)		Tareas propuestas por CRISP-DM + Ejecución de tareas de administración en un proceso aparte. Se sigue una guía de actividades para cumplir con las tareas. Técnicas propuestas
		Entendimiento de los datos	Uso del subproceso definido en CRISP-DM		Tareas propuestas por CRISP-DM + Técnicas propuestas
		Preparación de los datos	Uso del subproceso definido en CRISP-DM		Tareas propuestas por CRISP-DM + Se sigue una guía de actividades para cumplir con las tareas. Técnicas propuestas
		Modelado	Uso del subproceso definido en CRISP-DM		Tareas propuestas por CRISP-DM + Técnicas propuestas
		Evaluación	Uso del subproceso definido en CRISP-DM		Tareas propuestas por CRISP-DM + Se sigue una guía de actividades para cumplir con las tareas. Técnicas propuestas
Entrega	Uso del subproceso definido en CRISP-DM (incluye tareas propias de la administración mencionadas en el proceso correspondiente)		Tareas propuestas por CRISP-DM + Las tareas de administración se ejecutan en un proceso aparte. Se sigue una guía de actividades para cumplir con las tareas. Técnicas propuestas		

Con el uso del modelo propuesto podemos asegurar una mejora en el control de avance y performance de los proyectos en ejecución permitiendo que el éxito de un proyecto se base en el ajuste constante del proceso de desarrollo basado en experiencias pasadas.

Como futura línea de investigación y desarrollo se prevé trabajar en un análisis de los resultados de aplicación del modelo de procesos propuesto con el objetivo de evaluar ventajas y desventajas del mismo al momento de considerar su puesta en producción.

7. AGRADECIMIENTOS

Las investigaciones que se reportan en este artículo han sido financiadas parcialmente por el Proyecto de Investigación 33A105 del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús.

8. REFERENCIAS

[1] SEI 2006. CMMI® for Development, Version 1.2. Carnegie Mellon University, Software Engineering Institute.

[2] Oktaba, H. et al. 2005. Modelo de Procesos para la Industria de Software. Secretaría de Economía de México.

[3] Oktaba, H. et al. 2008. Competisoft, Mejora de Procesos Software para Pequeñas y Medianas Empresas y Proyectos. Ra-Ma.

[4] Vanrell, J. A. et al. 2010. Un Modelo de Procesos de Explotación de Información. Proceedings XII Workshop de Investigadores en Ciencias de la Computación, 167-171.

[5] García-Martínez, R. et al 2011. Towards and Information Mining Engineering. In Zapata, J. C. M. et al. (Eds.), Software Engineering, Methods, Modeling and Teaching, 83-99. Editorial Universidad de Medellín.

[6] Vanrell, J. A. 2009. Elementos para un Modelo de Procesos de Explotación de Información para PyMES. Trabajo de Especialidad en Ingeniería en Sistemas de Información. Escuela de Posgrado. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.

[7] Oktaba, H. et al. 2004. Método de Evaluación de Procesos para la Industria de Software. Secretaría de Economía de México.

[8] Chapman P. et al. 2000. CRISP-DM 1.0 Step-by-step data mining guide. The CRISP-DM consortium.

[9] Vanrell, J. A., et al. 2010. Modelo de Proceso de Operación para Proyectos de Explotación de Información. Anales del XVI Congreso Argentino de Ciencias de la Computación, 674-682.



[10] Britos, P. et al. 2008. Requiriments Elicitation in Data Mining for Business Inteligence Projects. In Avison, D. et al (Eds.), Advances in Information Systems Research, Education and Practice, 139-150. Springer.

[11] Flores, D. 2009. Detección de Patrones de Daños y Averías en la Industria Automotriz. Tesis de Maestría en Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.