

# UN MODELO DE PROCESOS PARA PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

Lic. Juan Angel Vanrell

Tesis de Maestría en Ingeniería en Sistemas de Información

Director: Dr. Ramón García-Martínez (UNLa)

Co-Director: Mg. Rodolfo Bertone (UNLP)

Escuela de Postgrado

Universidad Tecnológica Nacional

(FRBA)

2011

## Agradecimientos

A mi familia, mi mujer y mis amigos por darme el soporte para encarar todas las cosas de mi vida.

A mi director de tesis por la ayuda y la insistencia sin la cual esta tesis no sería posible.

## Índice

<b>1. Introducción.....</b>	<b>1</b>
1.1. Objetivos.....	1
1.1.1. Objetivo general.....	1
1.1.2. Objetivos específicos.....	2
1.2. Alcance.....	2
1.3. Fundamentos del trabajo.....	3
1.4. Metodología empleada.....	4
1.5. Producción científica vinculada a esta Tesis.....	5
1.6. Estructura de este trabajo.....	6
<b>2. Estado del Arte.....</b>	<b>9</b>
2.1. Generalidades.....	9
2.2. Competisoft.....	12
2.2.1. Alta Dirección.....	15
2.2.2. Gerencia.....	15
2.2.3. Operación.....	16
2.3. Explotación de Información.....	18
2.3.1. CRISP-DM.....	21
2.3.2. SEMMA.....	28
2.3.3. P <sup>3</sup> TQ.....	31
2.3.4. Comparación de las principales metodologías.....	38
2.3.5. Otras metodologías.....	39
2.3.6. Procesos de Explotación de Información.....	41
2.3.7. Roles en los Proyectos de Explotación de Información.....	42
2.4. Procesos de Competisoft cubiertos por las metodologías de explotación de información.....	43

---

<b>3. Problema</b> .....	<b>49</b>
3.1. Introducción.....	49
3.2. Descripción del Problema.....	50
3.3. Preguntas de Investigación.....	51
<b>4. Solución</b> .....	<b>53</b>
4.1. Propuesta.....	53
4.2. Solución basada en la propuesta.....	54
<b>5. Caso de Estudio</b> .....	<b>83</b>
5.1. Introducción.....	83
5.2. Comparación de uso.....	84
5.2.1. Preparación e inicio del proyecto.....	87
5.2.2. Ejecución del proyecto.....	89
5.2.3. Finalización del proyecto.....	92
<b>6. Conclusiones y Aportaciones de esta Tesis</b> .....	<b>93</b>
6.1. Conclusiones.....	93
6.1.1. Valoración sobre la investigación documental.....	93
6.1.2. Valoración del problema.....	94
6.1.3. Valoración de la solución.....	95
6.1.4. Valoración del Caso de Estudio.....	96
6.1.5. Respuestas a los interrogantes de investigación.....	97
6.2. Futuras líneas de trabajo.....	98
<b>7. Bibliografía</b> .....	<b>99</b>

## Índice de figuras

Figura 2.1	Diagrama de Categorías de Procesos.....	14
Figura 2.2	Relación entre los Elementos del Método de Evaluación ....	17
Figura 2.3	Niveles de Capacidad de Procesos.....	18
Figura 2.4	Fases del Modelo de Referencia CRISP-DM.....	22
Figura 2.5	Fases componentes de la Metodología CRISP-DM.....	24
Figura 2.6	Niveles de abstracción de Procesos de CRISP-DM.....	25
Figura 2.7	Fases de la Metodología SEMMA.....	28
Figura 2.8	Dinámica de la Metodología SEMMA.....	30
Figura 2.9	Fases de la Metodología P <sup>3</sup> TQ .....	37

## Índice de Tablas

Tabla 2.1	Tareas de cada Fase de la Metodología CRISP-DM.....	26
Tabla 2.2	Características de las principales Metodologías.....	39
Tabla 2.3	Procesos definidos en COMPETISOFT.....	45
Tabla 4.1	Proceso de Administración del Proyecto.....	59
Tabla 4.2	Proceso de Desarrollo del Proyecto.....	70
Tabla 5.1	Comparación de uso CRISP-DM vs. Modelo Propuesto.....	84

## Nomenclatura

CMM	Capability Maturity Model . Modelo de maduración de capacidades.
CMMi	Capability Maturity Model Integration.
Competisoft	Modelo de Procesos para el desarrollo de Software basado en MoProSoft.
CPM	Critical Path Method. Método para evaluación de camino crítico.
CRISP-DM	Metodología para el desarrollo de proyectos de explotación de información.
DIR	Categoría de Competisoft para la Alta Dirección.
DM-COMO	Data Mining Como. Modelo de estimación para proyectos de explotación de información.
EvalProSoft	Método de Evaluación de Procesos para la Industria de Software.
Gantt	Diagramas de que muestran el tiempo de dedicación previsto para diferentes tareas o actividades a lo largo de un tiempo total determinado.
GER	Categoría de Competisoft para la Gerencia.
ISO	International Organization for Standardization.
MEI	Modelo de Explotación de Información definido en P <sup>3</sup> TQ
MN	Modelo de Negocio definido en P <sup>3</sup> TQ.
MoProSoft	Modelo de Procesos para la Industria de Software desarrollado en Méjico.
mps Br	Melhoria do processo de software brasileiro. Proceso de mejoras para el desarrollo de software desarrollado en Brasil.
MUT	Método unificado de transformación.
OPE	Categoría de Competisoft para Operaciones.

---

P <sup>3</sup> TQ	Metodología para el desarrollo de proyectos de explotación de información (Product, Place, Price, Time, Quantity).
PERT	Project Evaluation and Review Technique. Técnica de evaluación y revisión de proyectos.
PyMEs	Pequeñas y medianas empresas.
RRHH	Recursos Humanos.
SAS	SAS Institute, Empresa creadora de SEMMA.
SEI	Software Engineering Institute.
SEMMA	Metodología para el desarrollo de proyectos de explotación de información (Sample, Expolore, Modify, Model and Assess).
SOM	Self-Organizing Map. Mapas auto organizados.
SPI	Software Process Improvement. Procesos de mejora de software.
SPIRE	Software Process Improvement in Regions of Europe. Procesos de mejora para el desarrollo de software en la región europea.
TDIDT	Top Down Induction of Decision Trees. Inducción de árboles de decisión.
TOPS	Toward Organised Software Process in SMEs.



## **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 propone la creación de un Modelo de Procesos para proyectos de Explotación de Información para PyMEs siguiendo los lineamientos del proyecto Competisoft.

## **Abstract**

Data Mining projects have very different features than classic software developer projects. Typical development phases like analysis, design, developing, integration and testing are not the natural phases of projects of Data Mining. In this context, we propose to create a Process Models for Data Mining projects for SEMs following the guidelines of the Competisoft Project.

# **1. Introducción**

En este Capítulo se presenta una introducción al trabajo que se realizó a lo largo de toda esta Tesis de Maestría, se mencionan los objetivos (sección 1.1), en particular el objetivo general (sección 1.1.2) y los objetivos específicos (sección 1.1.2), se define el alcance de este trabajo (sección 1.2), se evalúan los fundamentos que dan soporte a este trabajo (sección 1.3) y se describe la metodología empleada para el desarrollo del trabajo de Tesis (sección 1.4). El Capítulo finaliza con un resumen de las producciones científicas vinculadas con esta Tesis y con la descripción de la estructura del trabajo de la misma (sección 1.5).

## **1.1. Objetivos**

Dividimos los objetivos de esta Tesis en un objetivo general a alcanzar y un conjunto de objetivos específicos que definieron los pasos a seguir para lograr el objetivo general.

### **1.1.1. Objetivo general**

Se define como objetivo general de esta Tesis:

Construir un modelo de procesos para proyectos de explotación de información en Pymes (pequeñas y medianas empresas).

### **1.1.2. Objetivos específicos**

A continuación se detallan una serie de objetivos específicos que permiten en conjunto establecer los pasos a seguir para lograr cumplir con el objetivo general:

- Establecer diferencias entre los proyectos de Pymes y grandes empresas.
- Determinar límites, alcance y componentes del modelo Competisoft.
- Identificar los procesos específicos de los proyectos de explotación de información a través de las distintas metodologías existentes que se aplican a dichos proyectos.
- Seleccionar los procesos adecuados para el modelo a construir.
- Construir el modelo basándose en los procesos seleccionados.
- Evaluar la viabilidad de la propuesta.

## **1.2. Alcance**

El alcance de esta Tesis es establecer diferencias entre los proyectos para Pymes y grandes empresas justificando la creación de un modelo para Pymes en lugar de un modelo genérico.

Al mismo tiempo busca justificar el uso de modelos de procesos indicando las ventajas que acarreará su implementación en una empresa.

Se incluye una descripción del modelo Competisoft y de las distintas metodologías utilizadas en proyectos de explotación de información.

Finalmente se describe un modelo que incluye las carencias detectadas para proyectos de explotación de información y se evalúa su viabilidad realizando una comparación entre un proyecto desarrollado utilizando las metodologías tradicionales

y las diferencias que existirían de haberlo desarrollado utilizando el modelo propuesto.

### **1.3. Fundamentos del Trabajo**

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 [SEI, 2006], MoProSoft [Oktaba *et al.*, 2005] o su versión iberoamericana Competisoft [Oktaba *et al.*, 2008]. Todos estos son modelos genéricos por lo cual pueden ser utilizados para la ejecución de cualquier tipo de proyecto [Vanrell *et al.*, 2010b; García-Martínez *et al.*, 2011].

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 [Vanrell, 2009].

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.

## 1.4. Metodología empleada

Mediante este trabajo se pretende construir un modelo de procesos de explotación de información con el objetivo de facilitar el uso de modelos en pequeñas y medianas empresas que lleven a cabo proyectos de explotación de información.

En tal sentido, se describen los siguientes pasos metodológicos:

- La primera etapa consistió en la investigación y justificación de la distinción entre los modelos de desarrollo de productos de software aplicados a grandes empresas y los modelos aplicados a pequeñas y medianas empresas.
- En una segunda etapa se determinaron los límites, alcance y componentes del modelo Competisoft utilizando la documentación existente.
- Luego se investigaron las distintas metodologías existentes para los proyectos de explotación de información, utilizando la documentación de las mismas, se realizó una breve descripción de cada una de ellas y se hizo una comparación entre los distintos aspectos que abarca cada una.
- En una etapa posterior se establecieron procesos que se encuentran dentro del modelo de procesos Competisoft y se asociaron a los distintos procesos identificados en las metodologías descriptas en la etapa anterior.
- En el paso siguiente se identificaron los procesos que deben ser incluidos, excluidos o creados en el nuevo proceso para satisfacer las necesidades del sector.
- A continuación se construyó el modelo y se evaluó su viabilidad en la práctica.
- Finalmente se elaboró el informe final junto con las conclusiones obtenidas.

## 1.5. Producción científica vinculada a esta Tesis

Durante la construcción de esta Tesis se realizaron distintas comunicaciones de resultados parciales a través de un trabajo de especialización, publicaciones en congresos y artículos para revistas que se mencionan a continuación:

### Comunicaciones a Congresos:

Vanrell, J. A., Bertone, R., García-Martínez, R. (2010a). *Modelo de Proceso de Operación para Proyectos de Explotación de Información*. Anales del XVI Congreso Argentino de Ciencias de la Computación. Pág. 674-682. ISBN 978-950-9474-49-9.

Vanrell, J. A., García-Martínez, R., Bertone, R. (2010b). *Un Modelo de Procesos de Explotación de Información*. Proceedings XII Workshop de Investigadores en Ciencias de la Computación. Pág. 167-171. ISBN 978-950-34-0652-6.

### Artículos:

García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cataneo, F., Rodriguez, D., Pytel, P., Vanrell, J. A. (2011). *Towards and Information Mining Engineering*. Software Engineering, Methods, Modeling and Teaching. Pág. 83-99. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6.

### Trabajo de especialización:

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.

## **1.6. Estructura de este trabajo**

Este trabajo se compone de siete Capítulos que abarcan la totalidad del trabajo de Tesis, los cuales se detallan a continuación:

En el Capítulo 1 se presenta una introducción al trabajo que se realizó a lo largo de toda esta Tesis de Maestría, se establecen los objetivos, en particular el objetivo general y los objetivos específicos, se define el alcance de este trabajo, se evalúan los fundamentos que dan soporte a este trabajo y se describe la metodología empleada para el desarrollo del trabajo de tesis. El Capítulo finaliza resumiendo las producciones científicas vinculadas con esta Tesis y describiendo la estructura del trabajo de la misma.

En el Capítulo 2 se desarrollan generalidades relacionadas con la temática de la Tesis, a continuación se introduce el modelo de procesos Competisoft describiendo las distintas categorías en las que se divide el modelo, Alta Dirección, Gerencia y Operación. A continuación se realiza una descripción del área denominada Explotación de Información para luego pasar a describir las metodologías más utilizadas en este tipo de proyectos, CRISP-DM, SEMMA y P<sup>3</sup>TQ, luego se realiza una comparación entre todas estas metodologías para finalizar mencionando otras metodologías utilizadas, los procesos de Explotación de Información y los roles involucrados en los proyectos de este tipo. El Capítulo finaliza mencionando los procesos de Competisoft que son cubiertos por las metodologías de Explotación de Información como conclusión de las investigaciones realizadas.

En el Capítulo 3 se desarrolla la problemática que intenta solucionar esta Tesis junto con la justificación de las decisiones tomadas para llevar a cabo su resolución finalizando con las preguntas de investigación que se intentaron responder mediante este trabajo de tesis.

En el Capítulo 4 se desarrolla la solución propuesta para la problemática mencionada en el Capítulo anterior realizando una descripción de la solución planteada para luego dar lugar a las tablas que consisten en la solución propiamente dicha.

En el Capítulo 5 se describe el caso de estudio utilizado para evaluar la viabilidad de la propuesta, a continuación se realiza una comparación entre el uso de la metodología CRISP-DM en un caso real contra el uso del modelo propuesto como solución de la problemática planteada por esta Tesis la cual se divide en las diferencias y similitudes observadas en las tres etapas en las que se puede encontrar un proyecto, esto es, Preparación e Inicio del Proyecto, Ejecución del Proyecto y Finalización del Proyecto.

En el Capítulo 6 se describen las conclusiones obtenidas a partir del desarrollo de esta Tesis comenzando por las conclusiones que la justifican, realizando a continuación una valoración del problema encontrado, siguiendo con las conclusiones de la construcción de la solución y continuando con las conclusiones obtenidas en el caso de estudio utilizado para validar la propuesta, para finalizar se da respuesta a los interrogantes de investigación. Al final del capítulo se mencionan las futuras líneas de investigación a desarrollar a partir de esta Tesis.

El Capítulo 7 contiene las referencias bibliográficas utilizadas para este trabajo de tesis.





## 2. Estado del Arte

En este capítulo se desarrollan generalidades relacionadas con la temática de la Tesis (sección 2.1), a continuación se introduce el modelo de procesos Competisoft (sección 2.2) describiendo las distintas categorías en las que se divide el modelo, Alta Dirección (sección 2.1.1), Gerencia (sección 2.1.2) y Operación (sección 2.1.3). A continuación se realiza una descripción del área denominada Explotación de Información (sección 2.3) para luego pasar a describir las metodologías más utilizadas en este tipo de proyectos, CRISP-DM (sección 2.3.1), SEMMA (sección 2.3.2) y P<sup>3</sup>TQ (sección 2.3.3), luego se realiza una comparación entre todas estas metodologías (sección 2.3.4) para finalizar mencionando otras metodologías utilizadas (sección 2.3.5), los procesos de Explotación de Información (2.3.6) y los roles involucrados en los proyectos de este tipo (2.3.7). Al final de este capítulo se mencionan los procesos de Competisoft que son cubiertos por las metodologías de Explotación de Información como conclusión de las investigaciones realizadas (sección 2.4).

### 2.1. Generalidades

Mas y Amengual [2005] describen algunas características con las cuales se marca una diferencia entre las grandes empresas (de más de 200 desarrolladores) y las Pymes (menos de 30 desarrolladores). Estas características fueron divididas en categorías dentro de las cuales se identificaron distintos factores que diferencian a los grupos.

El primer factor es el de los recursos humanos, diferencia entre los tamaños de equipos y cantidad de los mismos, falta de roles especializados, responsabilidades no muy bien definidas y alta dependencia de los individuos ente otros. El segundo factor

identificado se relaciona con los aspectos económicos, en el caso de Pymes se da mayor importancia a la obtención de beneficios a corto plazo y la inversión en investigación y desarrollo suele ser mínima. Los procesos son identificados como un factor más en los cuales las Pymes encuentran serios problemas al intentar desarrollar y reflejar los resultados de la implantación de programas de mejora de procesos de software con el nivel de detalle y formalidad exigido por los grandes modelos. El último factor identificado tiene que ver con los proyectos, en donde las Pymes se destacan por trabajar en aquellos cuyo tamaño es reducido al igual que su duración, también se destacan en la cantidad de proyectos simultáneos que suelen ser pocos y las dificultades en alguno de los mismos suelen tener un alto grado de incidencia en la organización.

Los autores señalan algunos de los problemas que tiene las Pymes para adoptar grandes modelos de SPI como la duración promedio de los proyectos de SPI que van de 18 a 24 meses lo cual es mucho tiempo para una empresa de poca envergadura. Otro problema es el retorno de inversión que, si bien es alto (entre 4 y 9 veces el capital invertido), se produce a largo plazo. La implementación de dichos modelos de SPI encuentra entonces dificultades de aplicación en las Pymes principalmente debido a los costos de aplicación y al largo plazo para observar los resultados.

En el mismo sentido en [Oktaba *et al.*, 2007] se remarca la complejidad de las recomendaciones para la implementación de los grandes modelos y la implementación de los modelos construidos en otros países sin adaptación, coincidiendo con los factores de costo y tiempo antes mencionados.

Relacionado con el trabajo de SPI en Pymes el artículo [Pino *et al.*, 2006] indica que las pequeñas y medianas empresas son un engranaje muy importante en la economía mundial. En la mayoría de los países el desarrollo de software es llevado a cabo, en un gran porcentaje, por este tipo de empresas. Estas organizaciones,

denominadas Pymes\_DS, requieren prácticas eficientes de Ingeniería de Software adaptadas a su tamaño y tipo de negocio.

En la misma línea los autores recalcan que en la última década la comunidad vinculada a esta disciplina ha demostrado un gran interés en la mejora de procesos de software, buscando aumentar la calidad y productividad del software, lo cual se ve reflejado tanto en el creciente número de artículos sobre el tema como por la aparición de un gran número de iniciativas internacionales relacionadas con SPI.

Dentro de los trabajos e iniciativas a las que hacen referencia los autores para fortalecer SPI en Pymes\_DS podemos mencionar SPIRE (*Software Process Improvement in Regions of Europe*), TOPS (*Toward Organised Software Process in SMEs*) o el programa brasileiro PBQP-Software (*Productivity and Quality Software Program*) y el proyecto "mps Br" (*melhoria do processo de software brasileiro*). Puede sumárseles a estos trabajos e iniciativas el modelo mexicano MoProSoft [Oktaba *et al.*, 2005] (ver 3.1) y su iniciativa iberoamericana Competisoft [Oktaba *et al.*, 2008].

Entre las conclusiones que obtiene y citando a otros autores se encuentra que los estándares de facto (ISO y los modelos del SEI) difícilmente pueden ser aplicados a pequeñas empresas ya que "un proyecto de mejora supone una gran inversión de dinero, tiempo y recursos".

A continuación se realiza una descripción del modelo de procesos Competisoft que servirá como base para la construcción del nuevo modelo.

## 2.2. Competisoft

Competisoft [Oktaba *et al.*, 2008] se define como un proyecto de Mejora de Procesos para Fomentar la Competitividad de la Pequeña y Mediana Industria del Software de Iberoamérica. Se basa inicialmente en MoProSoft [Oktaba *et al.*, 2005] el cual es un modelo de procesos para el desarrollo de software 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, e incorporando el método de evaluación EvalProSoft [Oktaba *et al.*, 2004] y una división en etapas de madurez de los procesos similar a las identificadas en CMM o CMMI, entre otras pequeñas modificaciones.

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 o CMMI.

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 utilizarlo como punto de referencia para identificar los elementos que les hace falta cubrir.

Además de definir procesos los creadores de Competisoft definieron un patrón de trabajo que debe ser utilizado para documentar aquellos procedimientos 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

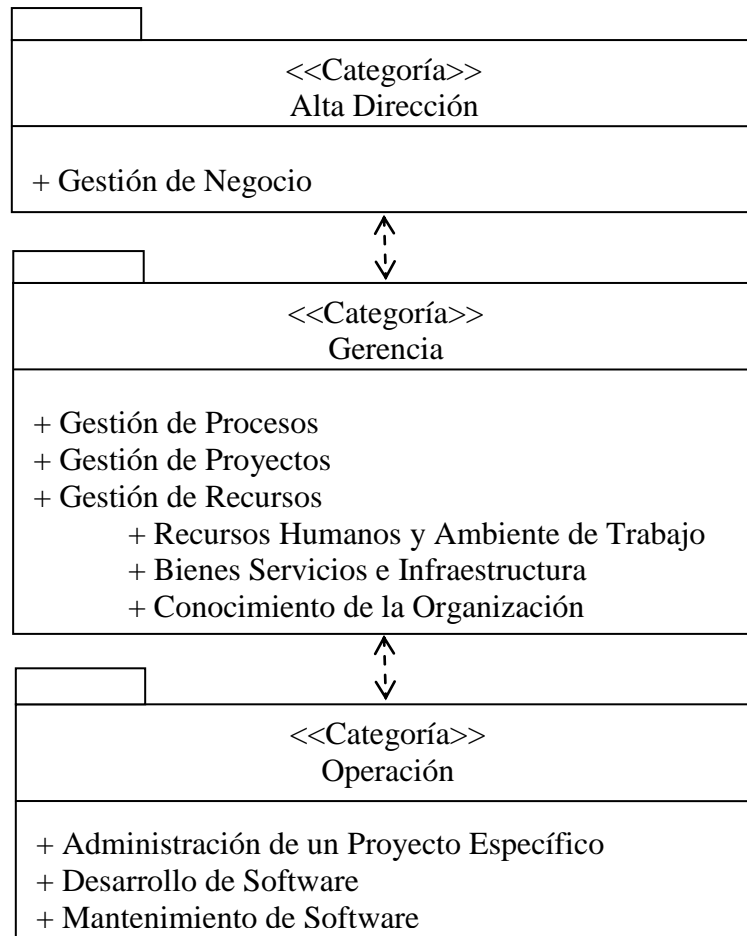
constituido por tres partes: Definición general del proceso, Prácticas y Guías de ajuste. Se seguirá este patrón para la documentación de los procesos de explotación de información que se incluirán en nuestro modelo.

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”.

Por último, la Categoría de Operación es la “categoría de procesos que aborda las prácticas de los proyectos de desarrollo de software 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”. Este esquema de categorías de procesos puede verse gráficamente en la *figura 2.1*.



**Figura 2.1:** Diagrama de categorías de procesos. Extraído de [Oktaba *et al.*, 2008]

### **2.2.1. Alta Dirección**

Dentro de la categoría de Alta Dirección se encuentra un único proceso, el de Gestión de Negocio. El propósito de este proceso es establecer la razón de ser de la organización, sus objetivos y las condiciones para lograrlos, para lo cual es necesario considerar las necesidades de los clientes, así como evaluar los resultados para poder proponer cambios que permitan la mejora continua.

También habilita a la organización para responder a un ambiente de cambio y a sus miembros para trabajar en función de los objetivos establecidos. Este proceso está compuesto por la planificación estratégica, la preparación para la realización de la estrategia, la valoración y mejora continua de la organización.

### **2.2.2. Gerencia**

Dentro de esta categoría se encuentran tres procesos, la Gestión de Procesos, la Gestión de Proyectos y la Gestión de Recursos. A su vez la Gestión de Recursos se divide en tres subprocesos, el subproceso de Recursos humanos y Ambiente de Trabajo, el de Bienes, Servicios e Infraestructura y el de Conocimiento de la Organización.

El proceso de Gestión de Procesos tiene como fin establecer los procesos de la organización en función de los procesos requeridos identificados en el plan estratégico. También incluye la definición, planificación e implantación de las actividades de mejora de los mismos. Incluye las actividades de planificación de procesos, la preparación para la implantación y la evaluación y control de procesos.

El propósito del proceso de Gestión de Proyectos es asegurar que los proyectos contribuyan al cumplimiento de los objetivos y estrategias de la organización. Sus actividades son la planificación, realización y evaluación de desempeño.



El proceso de Gestión de Recursos busca conseguir y dotar a la organización de los recursos humanos, infraestructura, ambiente de trabajo y proveedores, así como crear y mantener la base de conocimiento de la organización. Su finalidad es apoyar el cumplimiento de los objetivos del plan estratégico de la organización. Posee como actividades internas la planificación, seguimiento y control de recursos e investigación de tendencias tecnológicas, apoyadas con sus subprocesos.

El subproceso de Recursos Humanos y Ambiente de Trabajo es responsable de proporcionar los recursos humanos adecuados para cumplir las responsabilidades asignadas a los roles dentro de la organización, así como de la evaluación del ambiente de trabajo. Las actividades correspondientes a este subproceso incluyen la preparación instrumentación y generación de reportes.

El propósito del subproceso de Bienes, Servicios e Infraestructura es proporcionar proveedores de bienes, servicios e infraestructura que satisfagan los requisitos de adquisición de los procesos y proyectos. Las actividades que lo conforman son la preparación instrumentación y generación de reportes.

El fin del subproceso de Conocimiento de la Organización es mantener disponible y administrar la base de conocimiento que contiene la información y los productos generados por la organización. Sus actividades son la planificación, realización y seguimiento y control.

### **2.2.3. Operación**

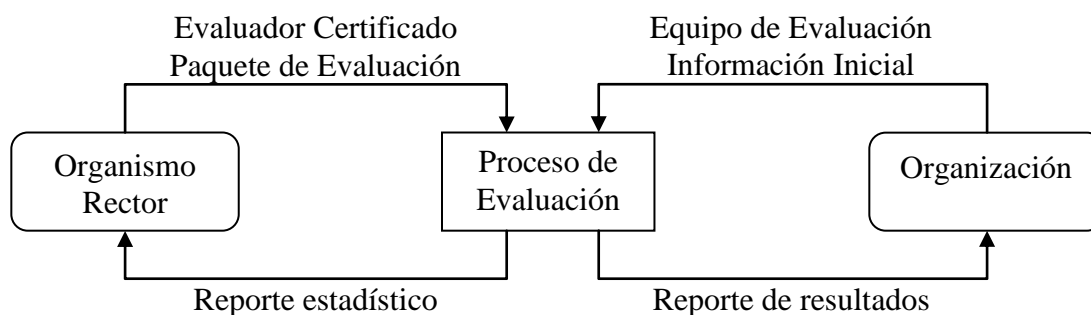
La categoría de operación incluye los procesos de Administración de Proyectos Específicos y el de Desarrollo y Mantenimiento de Software.

El proceso de Administración de Proyectos Específicos tiene como fin establecer y llevar a cabo sistemáticamente las actividades que permitan cumplir con

los objetivos de un proyecto en tiempo y costos esperados. Este proceso está compuesto por las actividades de planificación, realización, evaluación y control y cierre.

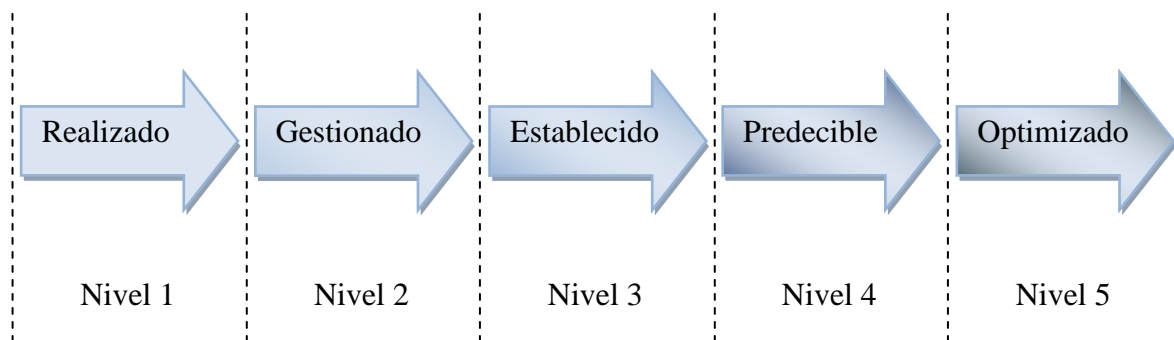
El proceso de Desarrollo de Software es la realización sistemática de las actividades de análisis, diseño, construcción, integración y pruebas de productos de software nuevos cumpliendo con los requisitos especificados y con las normativas de seguridad de la información. Este proceso está conformado por las actividades de inicio, requerimientos, análisis y diseño, construcción, integración y pruebas y cierre.

Competisoft cuenta además con un método de evaluación de los procesos empresariales para conformidad con el modelo desarrollado a partir de EvalProSoft [Oktaba, 2004]. Este método es utilizado para la acreditación de capacidades de la organización (a través de un evaluador certificado), para la evaluación de las capacidades de un proveedor (contratando un evaluador certificado que evalúe los procesos deseados del proveedor en cuestión) o para la autoevaluación de los procesos (no necesariamente con la utilización de un evaluador certificado). Los resultados de las evaluaciones son la base para el desarrollo de planes de mejora y son también utilizados para comparar el nivel de madurez de la organización con el de otras empresas, en el caso de evaluación de proveedores sirve para la selección del proveedor más adecuado. En la *figura 2.2* puede verse la relación entre los elementos del Método de Evaluación.



**Figura 2.2:** Relación entre los elementos del Método de Evaluación. Extraído de [Oktaba, 2004]

En EvalProSoft se define una categorización de niveles de capacidad de procesos que luego es incluida en Competisoft la cual es muy similar a la propuesta por CMM. Estos niveles pueden servir tanto para comparar la capacidad de los procesos de una empresa con respecto a otra así como para evaluar el nivel de madurez de una organización. En la *figura 2.3* pueden verse los niveles definidos. En el nivel 0 los procesos no se encuentran implantados o falta alcanzar el propósito completo del mismo. En un nivel 1 los procesos implantados logran su propósito. En nivel 2 los procesos realizados se implantan de forma administrada y sus productos de trabajo están apropiadamente establecidos, controlados y mantenidos. En el nivel 3 los procesos administrados son implantados mediante un proceso definido, el cual es capaz de lograr los resultados de los procesos. En un nivel 4 los procesos establecidos operan dentro de límites para lograr sus resultados. Por último, en el nivel 5 los procesos predecibles son continuamente mejorados para lograr las metas de negocio actuales y futuras relevantes.



**Figura 2.3:** Niveles de Capacidad de Procesos

## 2.3. Explotación de Información

Larose [2005] define el término Explotación de Información (*Data Mining*) como el proceso de descubrir nuevas correlaciones, patrones y tendencias utilizando grandes cantidades de datos almacenados en repositorios, usando tecnologías de

reconocimiento de patrones así como herramientas matemáticas y de estadística. Otras dos definiciones mencionadas en el mismo libro son el análisis de conjuntos de datos de observación para encontrar relaciones inesperadas y para resumir estos datos en nuevas formas que resulten útiles y entendibles para el dueño de los datos, o un campo interdisciplinario que junta técnicas de aprendizaje, reconocimiento de patrones, estadísticas, bases de datos y visualización para abordar problemas de extracción de información de grandes bases de datos.

En la referencia citada se destaca que el campo de explotación de información ha sido considerado como uno de los desarrollos revolucionarios de la próxima década y considerado como una de las 10 tecnologías emergentes que van a cambiar el mundo.

En [Aluja, 2001] el autor identifica los datos de una organización como un tesoro para las mismas dado que entre ellos se encuentran todas las interacciones pasadas con los clientes o la contabilidad de sus procesos internos. Estos datos representan la memoria de la organización pero tener memoria no es suficiente hay que pasar a la acción inteligente sobre esos datos para extraer la información que almacenan lo cual es el objetivo de la explotación de la información. Esta extracción de información permite abordar algunos problemas tales como: “Búsqueda de lo inesperado por descripción de la realidad multivariante”, “Búsqueda de asociaciones”, “Definición de tipologías”, “Detección de ciclos temporales” o “Predicción”.

En [Kuna *et al.*, 2009] se hace énfasis en el uso de técnicas, métodos y herramientas para garantizar, mediante un proceso formal, la calidad y seguridad de la información.

En la misma línea en [Sumathi y Sivanandam, 2006] se hace una descripción de los beneficios aportados por la utilización de un proceso general de explotación de información. Los beneficios señalados son:

- Un estándar puede incrementar el nivel de confort de clientes potenciales de los productos y servicios del proyecto de Explotación de Información.
- Una gran proporción del esfuerzo puede resultar en éxito dado que se siguen principios establecidos, incrementando la percepción del proyecto de Explotación de Información como una tecnología efectiva y de alto retorno de inversión.
- Los proyectos son más propensos a proceder sin sobresaltos si todas las partes entienden el curso normal de los proyectos de Explotación de Información.
- Un proceso estándar mejora la comunicación acerca de los tópicos del proyecto de Explotación de Información proveyendo un punto de referencia común para consumidores y clientes para comparar productos y servicios.

Rabuñal Dopico en [2002] indica algunas de las características de los enfoques de explotación de información presentando cinco características que son comunes a todos ellos. Estas características son:

- Requerimiento de grandes cantidades de datos para proporcionar información suficiente para derivar conocimiento adicional.
- Énfasis en el proceso de la eficiencia teniendo en cuenta la gran cantidad de datos.
- Exactitud para asegurar que el descubrimiento es válido.
- Presentación de los resultados de una manera entendible para el ser humano.

- El conocimiento debe ser descubierto utilizando técnicas de aprendizaje inteligente que van examinando datos a través de procesos automatizados.

Es claro que dada la cantidad de información a partir de la cual se trabaja es necesario contar con herramientas automatizadas que realicen el trabajo de la misma forma que será necesario seguir una serie de pasos preestablecidos para no cometer errores en el descubrimiento de la información, sobre todo teniendo en cuenta que a partir de los resultados se espera poder tomar decisiones que afectarán en mayor o menor grado a los dueños de la información.

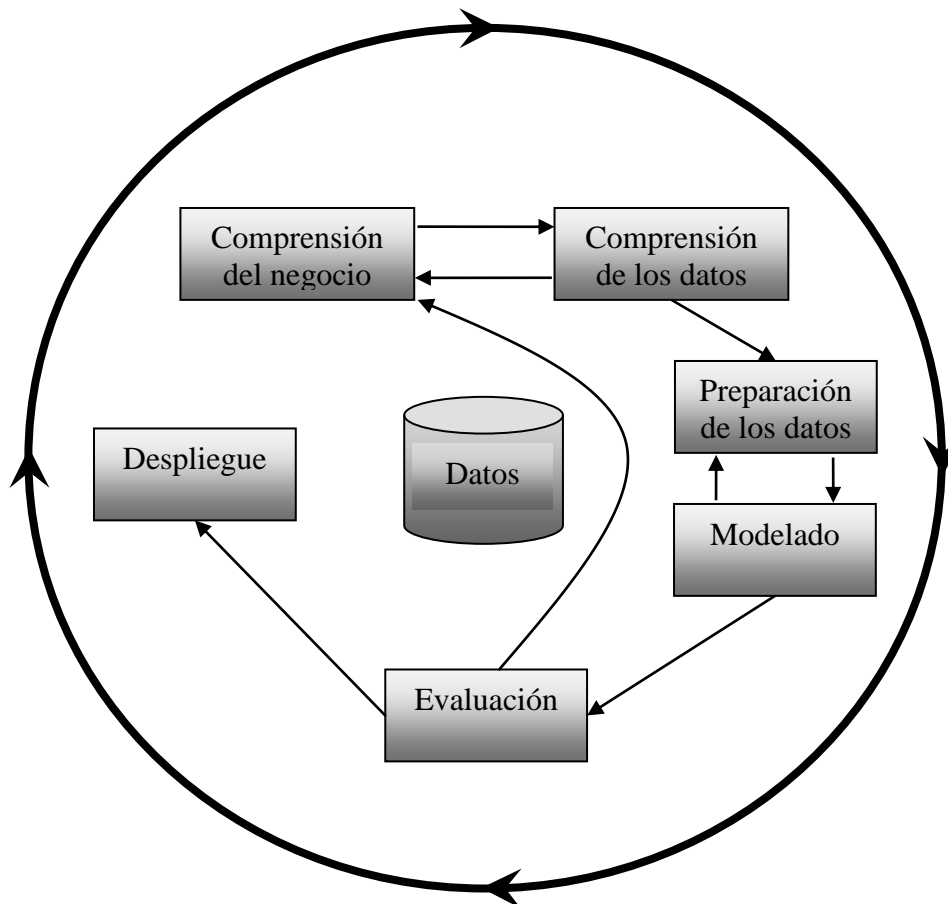
Existen actualmente varias metodologías de para proyectos de explotación de información, entre ellas podemos nombrar CRISP-DM, SEMMA y P<sup>3</sup>TQ como las más conocidas y algunas otras que no abarcan la totalidad de los proyectos sino que se enfocan en ciertos procesos de los mismos. Se utilizarán las distintas metodologías existentes para identificar procesos propios de este tipo de proyectos con el fin de incluirlos en el nuevo modelo.

### 2.3.1. CRISP-DM

La metodología CRISP-DM [Chapman *et al.*, 2000] se encuentra definida en base a un modelo jerárquico de procesos. Se mantendrá el foco 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 establece las principales fases de un proyecto de este tipo junto con las relaciones entre las mismas, como puede ser visto en la *figura 2.4*, estas relaciones son las más comunes aunque pueden establecerse entre cualquiera de las

fases. 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 describe cada una de las fases identificadas por CRISP-DM.



**Figura 2.4:** Fases del modelo de referencia CRISP-DM. Extraído de [Chapman *et al.*, 2000]

En la fase de Comprensión 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.

La Comprensión 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.

Durante la Evaluación se construyen varios modelos buscando aquellos que contribuyan a una mayor calidad de análisis, requiriendo la evaluación de cada modelo y la revisión de los pasos ejecutados para la construcción del mismo; a fin de asegurar el logro de los objetivos de negocio. La fase concluye al alcanzar una decisión en el uso 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.

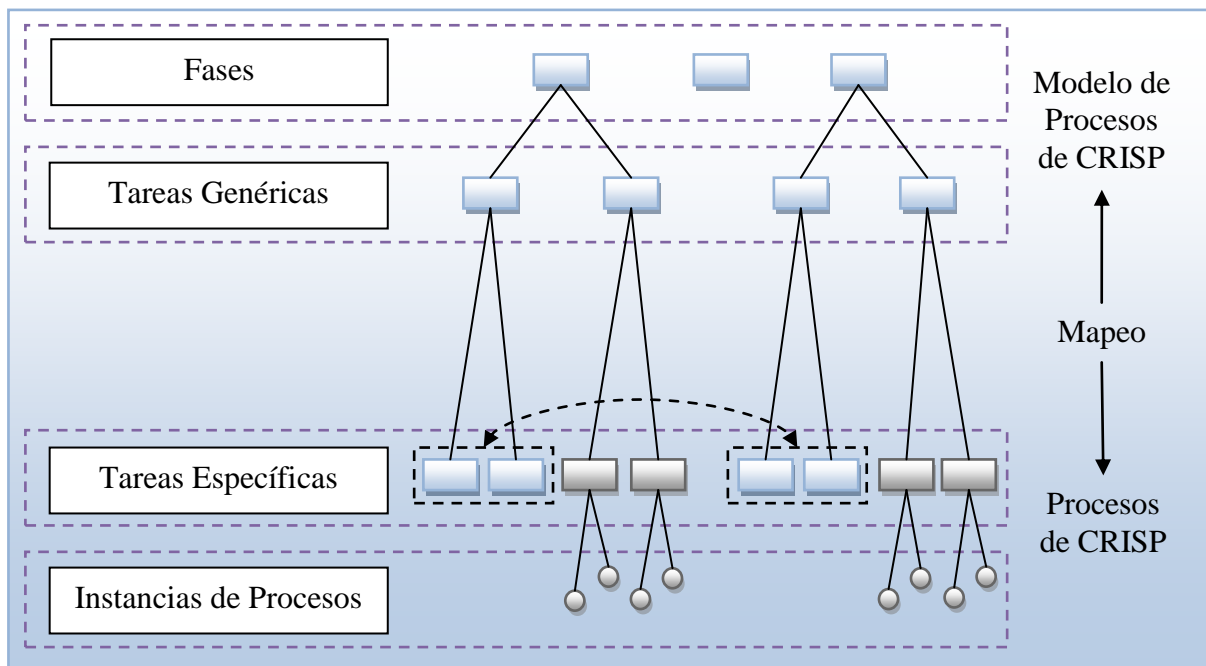
En la *figura 2.5* se observan claramente las fases que componen la metodología CRISP-DM en el orden secuencial natural de las mismas.





**Figura 2.5:** Fases componentes de la metodología CRISP-DM. Extraído de [Britos, 2008]

Cada una de estas fases se divide en distintas fases de nivel inferior que indican tareas generales a realizar dentro de la misma. A su vez, estas tareas de segundo nivel, son divididas en tareas específicas donde se describen las acciones que deben ser desarrolladas en situaciones específicas. Por ejemplo, si entre las tareas generales se encuentra “limpieza de datos” en el tercer nivel se indican las tareas que deben desarrollarse para un caso específico como por ejemplo “limpieza de datos numéricos” o “limpieza de datos categóricos”. En un cuarto nivel se recogen acciones, decisiones y resultados sobre el proyecto de Explotación de Información [Britos, 2008]. Esta abstracción de procesos puede verse gráficamente en la *figura 2.6* mientras que en la *tabla 2.1* pueden verse las tareas que corresponden a cada una de las fases.



**Figura 2.6:** Niveles de abstracción de procesos de CRISP-DM. Extraído de [Chapman *et al.*, 2000]

Para asistir al usuario de la metodología se proporcionan dos documentos distintos que serán utilizados como herramienta de ayuda al desarrollo de los proyectos de explotación de información: el modelo de referencia y la guía de usuario.

En el primer documento se describen en forma general las fases, tareas generales y salidas del proyecto de Explotación de Información en general.

La guía del usuario proporciona información más detallada sobre cómo aplicar el modelo de referencia a los proyectos específicos. Proporciona consejos y listas de comprobación sobre las tareas correspondientes a cada una de las fases.

**Tabla 2.1.a:** Tareas de cada fase de la metodología CRISP-DM. Extraído de [Britos, 2008]

FASE	TAREAS COMPONENTES	ACTIVIDADES ASOCIADAS
Comprensión del negocio	Determinar los objetivos del negocio	<ul style="list-style-type: none"> <li>▪ <i>Background</i></li> <li>▪ Objetivos del negocio</li> <li>▪ Criterios de éxito del negocio</li> </ul>
	Evaluar la situación	<ul style="list-style-type: none"> <li>▪ Inventario de recursos</li> <li>▪ Requisitos, supuestos y requerimientos</li> <li>▪ Riesgos y contingencias</li> <li>▪ Terminología</li> <li>▪ Costos y beneficios</li> </ul>
	Determinar objetivos del proyecto de Explotación de Información	<ul style="list-style-type: none"> <li>▪ Las metas del proyecto de Explotación de Información</li> <li>▪ Criterios de éxito del proyecto de Explotación de Información</li> </ul>
	Realizar el Plan de Proyecto	<ul style="list-style-type: none"> <li>▪ Plan de Proyecto</li> <li>▪ Valoración inicial de herramientas</li> </ul>
Comprensión de los datos	Recolectar los datos iniciales	<ul style="list-style-type: none"> <li>▪ Reporte de recolección de datos iniciales</li> </ul>
	Descubrir datos	<ul style="list-style-type: none"> <li>▪ Reporte de descripción de los datos</li> </ul>
	Explorar datos	<ul style="list-style-type: none"> <li>▪ Reporte de exploración de los datos</li> </ul>
	Verificar la calidad de los datos	<ul style="list-style-type: none"> <li>▪ Reporte de calidad de los datos</li> </ul>
	Seleccionar los datos	<ul style="list-style-type: none"> <li>▪ Inclusión / Exclusión de datos</li> </ul>

**Tabla 2.1.b:** Tareas de cada fase de la metodología CRISP-DM. Extraído de [Britos, 2008]

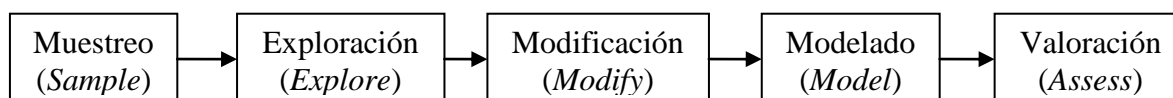
FASE	TAREAS COMPONENTES	ACTIVIDADES ASOCIADAS
Preparación de los datos	Caracterizar el conjunto de datos	<ul style="list-style-type: none"> <li>▪ Conjunto de datos</li> <li>▪ Descripción del conjunto de datos</li> </ul>
	Seleccionar los datos	<ul style="list-style-type: none"> <li>▪ Inclusión / Exclusión de datos</li> </ul>
	Limpiar los datos	<ul style="list-style-type: none"> <li>▪ Reporte de calidad de datos limpios</li> </ul>
	Estructurar los datos	<ul style="list-style-type: none"> <li>▪ Derivación de atributos</li> <li>▪ Generación de registros</li> </ul>
	Integrar los datos	<ul style="list-style-type: none"> <li>▪ Unificación de datos</li> </ul>
	Caracterizar el formato de los datos	<ul style="list-style-type: none"> <li>▪ Reporte de calidad de los datos</li> </ul>
	Generar el plan de pruebas	<ul style="list-style-type: none"> <li>▪ Plan de pruebas</li> </ul>
Modelado	Seleccionar una técnica de modelado	<ul style="list-style-type: none"> <li>▪ La técnica modelada</li> <li>▪ Supuestos del modelo</li> </ul>
	Generar el plan de pruebas	<ul style="list-style-type: none"> <li>▪ Plan de pruebas</li> </ul>
	Construir el modelo	<ul style="list-style-type: none"> <li>▪ Configuración de parámetros</li> <li>▪ Modelo</li> <li>▪ Descripción del modelo</li> </ul>
	Evaluar el modelo	<ul style="list-style-type: none"> <li>▪ Evaluar el modelo</li> <li>▪ Revisión de la configuración de parámetros</li> </ul>
Evaluación	Evaluar resultado	<ul style="list-style-type: none"> <li>▪ Valoración de los resultados mineros con respecto al éxito del negocio</li> <li>▪ Modelos aprobados</li> </ul>
	Revisar	<ul style="list-style-type: none"> <li>▪ Revisión del proceso</li> </ul>
	Determinar próximos pasos	<ul style="list-style-type: none"> <li>▪ Listar posibles acciones</li> </ul>
	Realizar la revisión del proyecto	<ul style="list-style-type: none"> <li>▪ Documentación de la experiencia</li> </ul>

**Tabla 3.c:** Tareas de cada fase de la metodología CRISP-DM. Extraído de [Britos, 2008]

FASE	TAREAS COMPONENTES	ACTIVIDADES ASOCIADAS
Implementación	Realizar el plan de implementación	<ul style="list-style-type: none"> <li>▪ Plan de implementación</li> </ul>
	Realizar el plan de monitoreo y mantenimiento	<ul style="list-style-type: none"> <li>▪ Plan de monitoreo y mantenimiento</li> </ul>
	Realizar el informe final	<ul style="list-style-type: none"> <li>▪ Informe final</li> <li>▪ Presentación final</li> </ul>
	Realizar la revisión del proyecto	<ul style="list-style-type: none"> <li>▪ Documentación de la experiencia</li> </ul>

### 2.3.2. SEMMA

Esta metodología define el proceso de selección, exploración y modelado de grandes cantidades de datos para descubrir patrones de datos desconocidos. Toma su nombre de las distintas etapas que conducen el proceso de explotación de información. En la *figura 2.7* se puede observar el ciclo de 5 etapas definidas por el instituto SAS para el proceso y la dinámica de la misma. SEMMA provee un proceso fácil de entender que permite el desarrollo y mantenimiento de proyectos de explotación de información organizado. [Britos, 2008; Azevedo *et al.*, 2008]

**Figura 2.7:** Fases de la metodología SEMMA. Extraído de [Britos, 2008]

Según [SAS, 1998] SEMMA es en sí misma un ciclo cuyos pasos internos pueden ser realizados iterativamente según las necesidades. En la *figura 2.8* puede verse como puede ser ejecutada la secuencia de pasos.

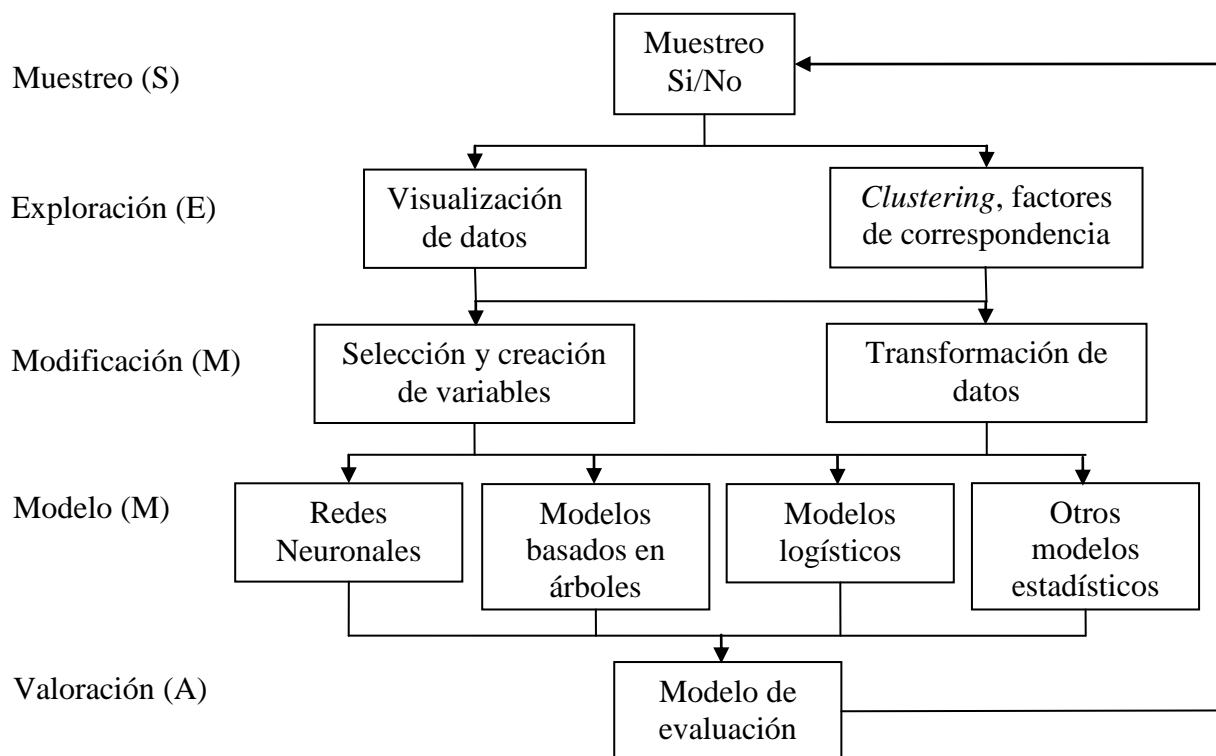
En el mismo artículo se menciona que aquellos proyectos que utilizan la metodología pueden evaluar millones de registros que permiten a las empresas

cumplir con los objetivos de la explotación de información como: segmentar adecuadamente grupos de clientes con patrones de compra similares, identificar perfiles de clientes para personalizar la gestión de relaciones, aumentar drásticamente la tasa de respuestas de campañas de correo, identificación de los clientes más rentables y su razón subyacente, identificar por que los clientes se cambian a la competencia, descubrir factores que afectan a la compra, etc.

Las etapas mencionadas son:

Etapa I: Muestreo (*Sample*) en la que se extrae la población muestral representativa sobre la cual se aplicará el análisis. La representatividad muestral es indispensable ya que de otro modo todo el modelo y los resultados dejan de ser fiables. La forma más común de seleccionar esta población es al azar, es decir, las probabilidades de selección de un individuo son exactamente las mismas para cada uno lo cual se denomina “muestreo aleatorio simple”. En la metodología cada una de las muestras se debe asociar a un nivel de confianza de la muestra.

Etapa II: Exploración (*Explore*) en donde se procede a realizar una exploración de la información para simplificar el problema y así optimizar la eficiencia del modelo. Para lograrlo se propone el uso de herramientas de visualización o técnicas estadísticas que ayuden a poner en manifiesto las relaciones entre las variables. Con esto se pretende determinar cuáles son las variables explicativas que se utilizarán como entradas del modelo.



**Figura 2.8:** Dinámica de la Metodología SEMMA. Extraído de [Britos, 2008]

Etapa III: *Modificación (Modify)* en la cual se modifican los datos de la base para que tengan el formato adecuado para la entrada del modelo.

Etapa IV: *Modelado (Model)* que permite modelar los datos permitiendo al software la búsqueda automática de una combinación de datos que predicen confiablemente las salidas deseadas. En esta etapa se debe establecer una relación entre las variables explicativas y las variables objeto del estudio, que posibiliten inferir el valor de las mismas con un nivel de confianza determinado. Las técnicas utilizadas para el modelado incluyen métodos estadísticos tradicionales, redes neuronales, técnicas adaptativas, lógica difusa, árboles de decisión, reglas de asociación y computación evolutiva.

Etapa V: *Valoración (Assess)* que consiste en la valoración de los datos evaluando usabilidad y confiabilidad de lo encontrado en el proceso y estimando que

tan bien se comporta haciendo una comparación con otros métodos estadísticos o con nuevas poblaciones muestrales.

### 2.3.3. P<sup>3</sup>TQ

La metodología P<sup>3</sup>TQ según [Britos, 2008] (Producto (*Product*), Lugar (*Place*), Precio (*Price*), Tiempo (*Time*) y Cantidad (*Quantity*)) está dividida en dos modelos, el Modelo de Negocio (MN) y el Modelo de Explotación de Información (MEI).

El primero de estos modelos “proporciona una guía de pasos para el desarrollo y la construcción de un modelo que permita identificar un problema de negocio o la oportunidad del mismo”, mientras que el segundo “proporciona una guía de pasos para la ejecución de modelos de Explotación de Información de acuerdo al modelo identificado en el (MN).

Ambos modelos poseen en su estructura los siguientes elementos:

- Una caja de actividades que indica una serie de pasos a realizar.
- Una caja de descubrimientos que provee acciones de exploración que se necesitan para poder decidir qué hacer en el próximo paso.
- Una caja de técnicas que proporciona información suplementaria sobre los pasos recomendados en las dos cajas anteriores.
- Una caja de ejemplos que dan una descripción detallada de cómo usar una técnica específica.

El modelado en (MN) depende de distintas circunstancias de negocio que promueven el planteo de 5 escenarios diferentes:

Escenario 1: Dato: el planteo comienza con una serie de datos y se debe explorar este conjunto para encontrar relaciones interesantes. En este caso se debe:



- 
- paso 1: Determinar la procedencia de los datos a recolectar.
  - paso 2: Identificar los recursos humanos para el proyecto.
  - paso 3: Discutir el proyecto con los recursos humanos.
  - paso 4: Caracterizar el conjunto de datos en término de las relaciones P<sup>3</sup>TQ (Producto, Lugar, Precio, Tiempo y Cantidad).
  - paso 5: Caracterizar la motivación del negocio para recolectar y almacenar los datos.
  - paso 6: Descubrir quienes o que departamento originó el proyecto y que se espera de él.

Escenario 2: Oportunidad: el planteo comienza con una situación de negocio, problema u oportunidad, que debe ser explorada. En este caso se debe:

- paso 1: Identificar las características de los recursos humanos relevantes.
- paso 2: Explotar las situaciones de negocio con los recursos humanos.
- paso 3: Determinar el marco de situación del negocio.
- paso 4: Definir los objetivos relevantes del negocio.
- paso 5: Buscar los datos a utilizar.
- paso 6: Presentar el caso de negocio a los recursos humanos.

Escenario 3: Prospectiva: el proyecto se diseña para descubrir donde la Explotación de Información puede ofrecer un valore en el entorno de la organización.

- paso 1: Caracterizar las claves de la organización en relación a P<sup>3</sup>TQ.
- paso 2: Identificar los principales procesos de flujo de información de la organización.
- paso 3: Identificar los potenciales recursos humanos.
- paso 4: Hablar con los potenciales recursos humanos.
- paso 5: Descubrir cuáles de los 26 niveles de gestión son los más involucrados para cada uno de los recursos humano.
- paso 6: Caracterizar los modelos más aplicables al negocio.
- paso 7: Explorar las fuentes de datos.
- paso 8: Preparar los casos de negocio para cada oportunidad significativa.

paso 9: Presentar el caso de negocio a los recursos humanos.

Escenario 4: Definido: el proyecto comienza con la premisa de crear la especificación del modelo de explotación con un propósito específico.

paso 1: Identificar los recursos humanos.

paso 2: Discutir los requerimientos con los recursos humanos.

paso 3: Enmarcar la situación de negocio.

paso 4: Buscar los datos necesarios.

paso 5: Definir los requerimientos a desarrollar.

Escenario 5: Estratégico: el proyecto comienza con una estrategia de análisis para dar soporte a un escenario planificado por la organización.

paso 1: Identificar los recursos humanos potenciales.

paso 2: Hablar con los recursos humanos potenciales.

paso 3: Enmarcar la situación de negocio.

paso 4: Si es necesario, trabajar interactivamente con los recursos humanos para crear un mapa de los escenarios estratégicos.

paso 5: A partir del mapa, crear un modelo sistémico de la situación estratégica.

paso 6: Caracterizar las claves de la organización en relación a P<sup>3</sup>TQ.

paso 7: Relacionar el mapa con las claves de la organización en relación a P<sup>3</sup>TQ.

paso 8: Si es necesario, simular una situación estratégica para descubrir incertidumbres, ambigüedades, errores en la comprensión intuitiva y descubrir relaciones cruciales (aquellas no entendidas, bien caracterizadas, o para las cuales los pequeños cambios tienen grandes efectos).

paso 9: Caracterizar las claves de las relaciones de los sistemas en término de los 26 niveles de gestión.

paso 10: Descubrir cuáles de los 26 niveles de gestión son los más involucrados para cada recurso humano.

paso 11: Caracterizar los modelos de los niveles de negocio más aplicados.

paso 12: Explorar las fuentes de datos.

paso 13: Enmarcar cada problema de negocio u oportunidad en un modelo estratégico con particular atención en las estrategias, interacciones estratégicas y los riesgos incluidos en el banco de pruebas de riesgo y las expectativas.

paso 14: Explotar los datos para caracterizar las relaciones actuales con el sistema modelado y la simulación, tratando de que concuerden con la situación real.

paso 15: Mostrar las relaciones descubiertas dentro de los mapas del sistema y la simulación y realizar la simulación a través del rango de escenarios requeridos.

Una vez definido el escenario se toman en cuenta distintos parámetros concernientes a la organización y a la situación del proyecto. Estos parámetros son tratados de distinta manera según el punto de partida, para obtener finalmente los datos requeridos para el proyecto de explotación de información y los requerimientos reales de las partes interesadas.

Para el modelado en (MEI) se siguen los pasos:

Paso 1: Preparación de los datos, en el cual se debe:

- Comprobar las variables de la matriz de característica.
- Comprobar las variables básicas para el problema.
- Comprobar los datos básicos para el problema.
- Comprobar las variables anacrónicas.
- Comprobar la suficiencia de los datos.
- Comprobar la representación de los resultados.
- Comprobar la representación de rasgos básica.

---

Paso 2: Selección de herramientas y modelado inicial, en el cual se debe:

- Definir la estructura de datos para llevar adelante la Explotación de Información.
- Caracterizar los datos de entradas y salidas.
- Seleccionar las herramientas de Explotación de Información.
- Construir los valores que comprueben el modelo.
- Si los datos no se comprenden: Crear el modelo exploratorio inicial.
- Si se van a clasificar los datos: Descubrir el tipo apropiado de modelo de clasificación inicial.
- Si se van a predecir los datos: Descubrir el tipo apropiado de modelo predictivo.

Paso 3: Ejecución, en el cual:

- Si es un modelo deductivo: Especificar la explicación del mismo.
- Si el modelo de clasificación o predicción es binario: Especificar una matriz de confusión.
- Si el modelo de clasificación o predicción es un valor continuo: Especificar una matriz de confusión, comparar la predicción con un grafico residual, comparar la predicción con la situación actual.
- Si el modelo de clasificación o predicción es una clase: Especificar una matriz de confusión, comparar la predicción con un grafico residual, comparar la predicción con un argumento actual, especificar pruebas del modelo residual.
- Si el modelo de clasificación o predicción es un valor categórico: Especificar la predicción con un grafico residual, comparar la predicción con situaciones actuales, especificar pruebas del modelo residual, realizar histogramas residuales, comparar situaciones

actuales con gráficos residuales XY, comparar la situación actual con una predicción de rangos, comparar la situación actual con curvas de predicción, comparar la situación actual con la predicción apta, especificar la varianza residual, especificar el modelo perfecto.

- apta, especificar la varianza residual, especificar el modelo perfecto.

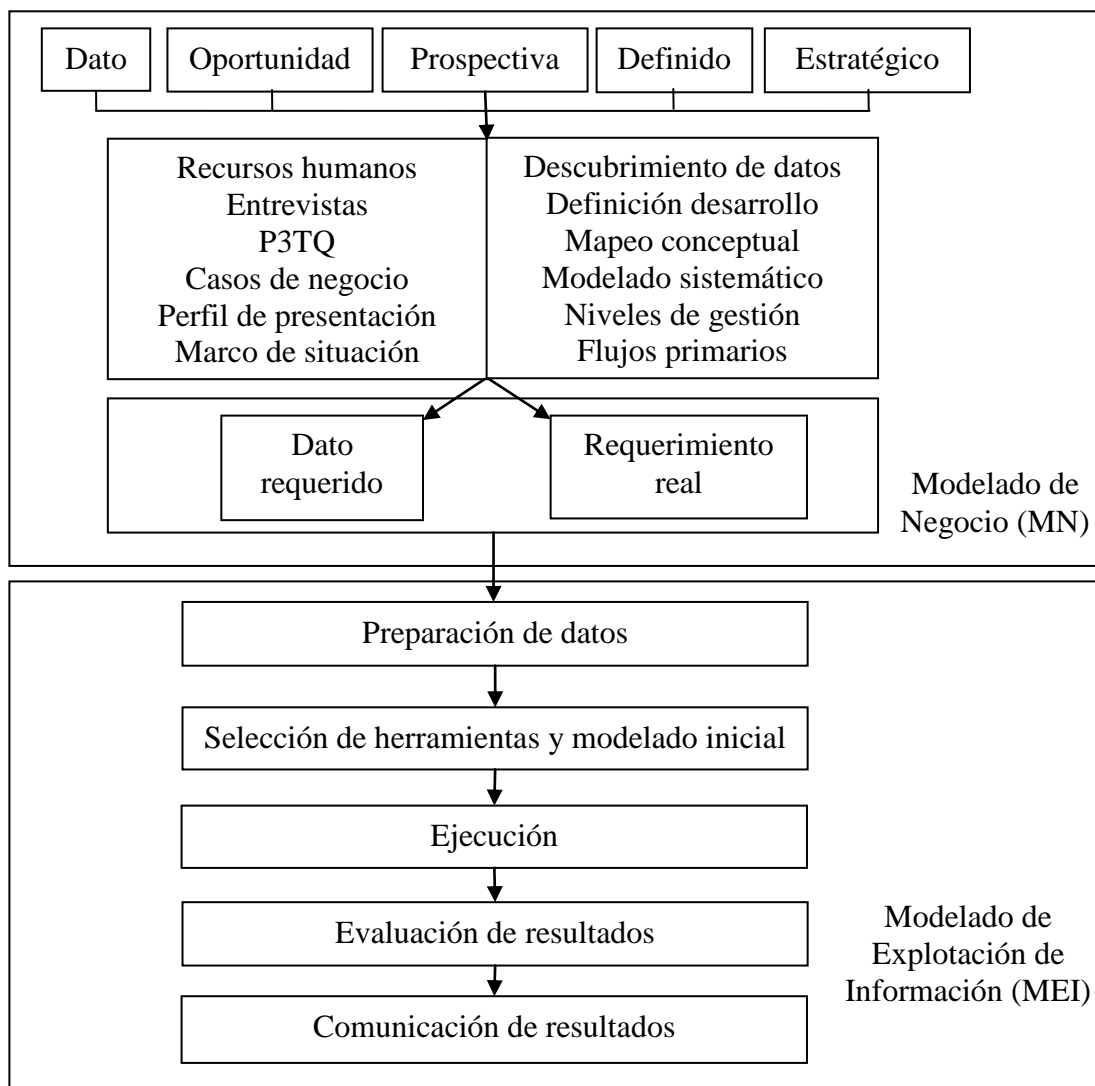
Paso 4: Evaluación de resultados, en el cual:

- Si es un modelo deductivo: Revisar los requerimientos descubiertos durante la ejecución, explicar en forma narrativa: a) los descubrimientos, el informe debe incluir: patrones, descubrimiento de explicaciones plausibles, clustering, conteos, contrastes y comparaciones, variables de particionamiento, generalidades de particularidades, proponer factores plausibles explícitos e implícitos latentes, identificar y explicar las relaciones entre variables (o variables grupales), crear explicaciones de cambios lógicos, creando coherencias conceptuales; y b) la verificación, el informe debe incluir: comprobación de la representatividad, comprobación de la tendencia, triangulación (usando fuentes de datos diferentes, usando métodos de modelado diferentes, utilizando teorías diferentes), considerando los límites, incorporando pruebas negativas, incorporando pruebas externas empíricas.
- Si es un modelo de clasificación: Revisar las exigencias de la entrega desarrollada antes de la ejecución del proceso de Explotación de Información, repasar los descubrimientos realizados durante la formación, preparar una explicación de soporte, crear la calibración de los modelos, revisar los modelos requeridos para entregar.
- Si el modelo es en tiempo real: Identificar las novedades.

Paso 5: Comunicación de resultados, en el cual se debe:

- Dar a las partes restantes del proyecto los resultados y sugerir como implementarlos.

En la *figura 2.9* puede verse cómo interactúan los modelos y la secuencia de pasos.



**Figura 2.9:** Fases de la metodología P3TQ. Extraído de [Britos, 2008]

Cada una de las partes de la metodología (MN y MEI) se encuentra desagregada en pasos denominados boxes, existen tres tipos distintos de ellos:

- *Action Boxes*: en donde se decide cuál es el próximo paso a realizar.
- *Discovery Boxes*: en donde se analizan los posibles resultados y problemas luego de ejecutar un *Action Box*.
- *Technique Boxes*: las cuales describen minuciosamente cómo debe emplearse una técnica.

Estos boxes no son recorridos secuencialmente sino que los saltos entre ellos dependen de las situaciones que se van sucediendo a medida que avanza el proyecto. Éstos explican detalladamente los conceptos y/o acciones que se realizan [Mendez, 2009].

#### **2.3.4. Comparación de las principales metodologías**

Mendez [2009] realiza una comparativa entre las principales metodologías utilizadas para los proyectos de explotación de información, sus conclusiones pueden visualizarse en la *tabla 2.2*, en estas se destaca la categorización de SEMMA como una metodología centrada en los aspectos técnicos del proyecto además de ser dependiente de los productos desarrollados por SAS para los cuales fue desarrollada. Destaca de CRISP-DM el ser más completa y abierta que SEMMA pero pone en evidencia la falta de detalle a la que llega P<sup>3</sup>TQ al nombrar etapas del proceso sin analizar los pasos, resultados y situaciones que se pueden dar dentro de cada capa.

A estas conclusiones pueden añadirse las obtenidas por Britos [2008] entre las cuales se menciona que, si bien las tres metodologías identifican técnicas de explotación de información utilizables, solo CRISP-DM identifica los distintos problemas de inteligencia de negocio y hace una caracterización parcialmente abstracta de los mismos.

Al mismo tiempo, CRISP-DM identifica relaciones entre las técnicas de explotación de información y las variables que modelan los problemas de inteligencia de negocio esbozando parcialmente los procesos a desarrollar. Esto nuevamente no ocurre en las otras dos metodologías.

**Tabla 2.2:** Características de las principales metodologías. Extraído de [Mendez, 2009] y [Britos, 2008]

	SEMMA	CRISP-DM	P <sup>3</sup> TQ
Permite elección libre de las herramientas	NO	SI	SI
Cantidad de fases	5	6	5 (1 MN Y 4 MEI)
Todas las fases pueden relacionarse	NO	SI	SI
Considera los motivos del proyecto	NO	NO	SI
Considera la naturaleza del interés de las partes	NO	NO	SI
Considera otros aspectos no técnicos	NO	SI	SI
Identifica claramente las variables sobre las cuales el proyecto tiene impacto	NO	NO	SI (Producto, Lugar, Precio, Tiempo, Cantidad)
Está detallada paso a paso cada etapa del método	NO	NO	SI
Identifica problemas de inteligencia de negocio (PIN)	NO	SI	NO
Identifica una caracterización abstracta de PIN	Parcialmente	NO	NO
Identifica técnicas de Explotación de Información (TEI) utilizables	SI	SI	SI
Identifica relaciones entre las TEI y los PIN	Parcialmente	NO	NO
Identifica procesos de explotación de información (procesos PINxTEI)	Parcialmente	NO	NO

### 2.3.5. Otras metodologías

Las metodologías antes mencionadas son las más utilizadas en la actualidad, éstas abarcan la mayor parte de los procesos necesarios para llevar a cabo exitosamente un proyecto de explotación de información. Sin embargo existen otras metodologías o trabajos en el área que las complementan.



En [Ochoa, 2006], por ejemplo, se propone una profase que sirve como complemento a cualquier metodología, ya sea de explotación de información o de otro tipo de proyecto, en la cual se persigue la comprensión clara del dominio de negocio. Este trabajo incrementa las fases que deben ser consideradas en cualquier tipo de proyecto, y en particular en los de explotación de información, de forma que al inicio del proyecto se realice una comprensión del entorno.

En este sentido se explica que todo proyecto se inicia con un problema a resolver y dicho problema existe en un entorno específico que condiciona los resultados y soluciones. Señala además que un inapropiado entendimiento del negocio puede hacer fracasar cualquier proyecto de software y que su comprensión no solo ayuda a determinar el tipo de solución a proponer sino que además la condiciona.

Para ello propone una serie de técnicas y herramientas que ayudan a la comprensión del negocio.

A nuestro entender, la fase propuesta, debe ser incluida entre los procesos de explotación de información para mejorar la calidad de los mismos y minimizar el impacto negativo que pueda acarrear la falta de conocimiento de negocio.

En [Amón y Jiménez, 2009] se propone la creación de una metodología para la selección de técnicas de depuración. Es común poseer, en los repositorios de información, errores de digitación, datos inconsistentes, valores ausentes o duplicados, estos problemas deterioran la calidad de los datos y en consecuencia la calidad de las decisiones que se tomen basándose en ellos. Los autores justifican la creación de una nueva metodología al considerar que las metodologías más amplias como CRISP-DM o SEMMA no presta suficiente detalle a la transformación y limpieza de datos.

### 2.3.6. Procesos de Explotación de Información

A partir del trabajo realizado por [Britos, 2008] se proponen cinco procesos de explotación de información que pueden ser considerados por CRISP-DM dentro de la etapa de Modelado, en SEMMA en la etapa de Modelo y en P<sup>3</sup>TQ en la etapa de Identificación de Procesos y Modelado Inicial.

Los procesos de explotación de información definidos son los siguientes:

- Descubrimiento de Reglas de Comportamiento
- Descubrimiento de Grupos
- Ponderación de Interdependencia de Atributos
- Descubrimiento de Reglas de Pertenencia a Grupos
- Ponderación de Reglas de Comportamiento o de la Pertenencia a Grupos

El proceso de Descubrimiento de Reglas se utiliza al querer identificar condiciones para obtener resultados del dominio del problema. Puede ser utilizado para descubrir las características del local más visitado por los clientes o establecer las características de los clientes con alto grado de fidelidad a la marca.

El proceso de Descubrimiento de Grupos es útil en los casos en que necesitemos identificar una partición dentro de la información disponible dentro del dominio de un problema. Como ejemplos de este tipo de procesos Britos menciona la identificación de tipos de llamadas que realizan los clientes de una empresa de telecomunicaciones o la identificación de grupos sociales con las mismas características entre otros.

El proceso de Ponderación de Interdependencia de Atributos se utiliza cuando se desea identificar los factores con mayor incidencia sobre un determinado resultado de un problema. Son ejemplos aplicables a este proceso la determinación de factores que

poseen incidencia sobre las ventas o la individualización de atributos clave que convierten en vendible a un determinado producto.

El proceso de Descubrimiento de Reglas de Pertenencia a Grupos es utilizado cuando se necesita identificar las condiciones de pertenencia a cada una de las clases en una partición desconocida pero que se encuentra presente en la masa de información disponible sobre el dominio del problema. Este tipo de proceso puede ser utilizado para la segmentación etaria de estudiantes y el comportamiento de cada segmento o la determinación de las clases de las llamadas telefónicas en una región y caracterización de cada clase.

Por último, el proceso de Ponderación de Reglas de Comportamiento de la Pertenencia a Grupos se utiliza cuando se requiere identificar las condiciones con mayor incidencia sobre la obtención de un determinado resultado en el dominio del problema, ya sea por la mayor medida en la que inciden sobre su comportamiento o las que mejor definen la pertenencia a un grupo. Como ejemplos de este tipo de proceso se puede citar la identificación del factor dominante que incide en el alza de ventas de un producto dado o el rasgo con mayor presencia en los clientes con alto grado de fidelidad a la marca.

En su trabajo Britos propone distintas tecnologías que pueden ser aplicadas a cada uno de los procesos sin embargo estos procesos son independientes de la tecnología que se utiliza para resolverlos.

### **2.3.7. Roles en los Proyectos de Explotación de Información**

Los proyectos de explotación de información poseen roles bien definidos para las personas que intervienen en un proyecto. En este sentido [Jackson, 2002] identifica los distintos actores en explotación de información como:

- Rol 1: *Líder de Proyecto*: es aquél que tiene la responsabilidad de planear, coordinar, ejecutar e implementar el proyecto.
- Rol 2: *Cliente del proyecto de Explotación de Información*: es el experto en el dominio que requiere el proyecto y que utilizará los resultados pero que no posee generalmente los conocimientos requeridos para participar en la ejecución de las fases más técnicas del proyecto como la preparación de los datos o el modelado.
- Rol 3: *Analista de Explotación de Información*: es quien posee una gran comprensión, desde la perspectiva de negocio, de lo que el cliente desea lograr y asiste en la traducción de estos objetivos en requerimientos técnicos a ser utilizados para la construcción de modelos.
- Rol 4: *Ingeniero en Explotación de Información*: es aquel que desarrolla, interpreta y evalúa los modelos de *Explotación de Información* en base a los objetivos de negocio y criterios de éxito, realiza las tareas en constante consulta con el cliente y el analista para ser asistido en el logro del fin de negocio.
- Rol 5: *Analista IT*: su responsabilidad es proveer el acceso al hardware, software y datos necesarios para completar el proyecto exitosamente).

Dependiendo de la envergadura del proyecto estos roles pueden ser asumidos por varios individuos o un mismo individuo asumir varios de ellos.

## **2.4. Procesos de Competisoft cubiertos por las metodologías de explotación de información**

En la *tabla 3* se presenta un resumen de los procesos existentes en Competisoft junto con los procesos identificados en cada una de las metodologías principales

(CRISP-DM, SEMMA y P<sup>3</sup>TQ) que cubren en forma total o parcial las tareas a realizar.

A partir del resumen presentado se pueden obtener distintas conclusiones.

El proceso de “Gestión de Negocio” dentro de la categoría de procesos de “Alta Dirección” no es considerado en ninguna de las metodologías. Sin embargo este proceso contempla actividades comunes para cualquier tipo de negocio, por lo cual puede ser tomado sin modificaciones para un modelo de procesos de proyectos de explotación de información.

Los procesos “Gestión de Procesos” y “Gestión de Proyectos” dentro de la categoría de “Gerencia” no se encuentran contemplados por ninguna de las metodologías. Nuevamente estos procesos son lo suficientemente genéricos como para ser utilizados en cualquier organización independientemente del tipo de proyectos que esté realizando, con lo cual pueden ser utilizados sin modificaciones.

En el caso del proceso de “Gestión de Recursos” se presentan distintas alternativas dependiendo del subproceso seleccionado y de la metodología a evaluar.

En el caso del subproceso de “Recursos Humanos y Ambiente de Trabajo” tanto CRISP-DM como SEMMA no poseen procesos o tareas relacionadas. Sin embargo la metodología P<sup>3</sup>TQ posee una tarea identificada como “Identificación de los recursos humanos” que se encuentra dentro del “Modelado del Negocio” que tiene alta relación con el mismo. Este proceso debería ser adecuado para contemplar la tarea mencionada en el caso de proyectos de explotación de información.

**Tabla 2.3:** Procesos definidos en Competisoft y su relación con las distintas metodologías.

Competisoft		CRISP-DM	SEMMA	P <sup>3</sup> TQ
Gestión de Negocios		No contemplado	No contemplado	No contemplado
Gestión de Procesos		No contemplado	No contemplado	No contemplado
Gestión de Proyectos		No contemplado	No contemplado	No contemplado
Gestión de Recursos	RRHH y Ambiente de Trabajo	No contemplado	No contemplado	Modelado del Negocio (Identificación de los recursos humanos)
	Bienes y Servicios e Infraestructura	Inventario de Recursos	No contemplado	No contemplado
	Conocimiento de la organización	Determinación de los objetivos de negocio  Evaluación de la situación  Determinación de las metas del proyecto de Explotación de Información	No contemplado	Modelado del Negocio
Administración de Proyectos Específicos		Producción del Plan de Proyecto	No contemplado	No contemplado
Desarrollo de Software		Entendimiento de los datos  Preparación de los datos  Evaluación  Modelado  Despliegue	Muestreo  Exploración  Modificación  Modelado  Valoración	Preparación de datos  Selección de herramientas y modelado inicial  Ejecución  Evaluación de resultados  Comunicación de resultados

Para el subproceso de “Bienes y Servicios e Infraestructura” ni SEMMA ni P<sup>3</sup>TQ poseen tareas relacionadas. En este caso es CRISP-DM la metodología que aporta elementos para este proceso en la tarea identificada como “Inventario de Recursos”. Nuevamente en este caso el proceso deberá ser adecuado para contemplar los cambios necesarios para soportar esta tarea.

El subprocesos de “Conocimiento de la Organización” no es contemplado por SEMMA pero si por CRISP-DM al igual que por P<sup>3</sup>TQ. En el caso de CRISP-DM se definen las tareas “Determinación de los objetivos”, “Evaluación de la situación” y “Determinación de las metas del proyecto de Explotación de Información” las cuales se encuentran estrechamente relacionadas con el proceso. P<sup>3</sup>TQ define una evaluación completa del negocio denominada “Modelado del Negocio” que puede ser contemplada en el proceso de “Conocimiento de la Organización” en un proyecto de explotación de información. Este proceso deberá ser adecuado para contemplar las tareas definidas en las metodologías mencionadas.

Dentro de la categoría de “Operación” el proceso “Administración de Proyectos Específicos” no se encuentra contemplado por las metodologías SEMMA o P<sup>3</sup>TQ. CRISP-DM incluye la tarea denominada “Producción del Plan de Proyecto”. Esta tarea debería ser contemplada en el proceso para los proyectos de explotación de información.

El proceso “Desarrollo de Software” dentro de la misma categoría es contemplado por las tres metodologías lo cual demuestra que las metodologías de proyectos de explotación de información se encuentran altamente vinculadas al desarrollo más que al gerenciamiento de los proyectos.

Puntualmente SEMMA es una metodología completamente orientada a este proceso dado que todos sus elementos son vinculados con el mismo. Los elementos mencionados son las tareas de Muestreo, Exploración, Modificación, Modelado y Valoración. En el caso de CRISP-DM las tareas relacionadas con este proceso son las de Entendimiento de los datos, Preparación de los datos, Evaluación, Modelado y Despliegue. Por último para P<sup>3</sup>TQ las tareas relacionadas con este proceso son Preparación de datos, Selección de herramientas y modelado inicial, Ejecución, Evaluación de resultados y Comunicación de resultados.

Es claro que este proceso difiere completamente en el caso de proyectos de explotación de información por lo cual debe ser completamente redefinido para adecuarlo a este tipo de proyectos.





## 3. Problema

En este capítulo se desarrolla la problemática que intenta solucionar esta Tesis (sección 3.1) junto con la justificación de las decisiones tomadas para llevar a cabo su resolución (sección 3.2) finalizando con las preguntas de investigación que se intenta responder mediante este trabajo de tesis (sección 3.3).

### 3.1. Introducción

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 si 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, P<sup>3</sup>TQ 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 los proyectos y de empresa.

## 3.2. Descripción del Problema

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 [Vanrell *et al.*, 2010a], 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, entre otras.

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 esta Tesis se utiliza como metodología de referencia a CRISP-DM dado que, cuando se compararon las tres metodologías mencionadas anteriormente [Britos *et al.*, 2008], se pudo observar que esta última contenía todos los elementos a nivel operación de las otras.

### **3.3. Preguntas de Investigación**

Teniendo en cuenta el problema planteado surgen los siguientes interrogantes que se intentan responder a lo largo de este trabajo de tesis:

- ¿Es posible construir un modelo de procesos para proyectos de Explotación de Información que posea procesos bien diferenciados para desarrollo y gestión?
- ¿Pueden ser adecuados los procesos de los modelos de desarrollo de software tradicional para los procesos necesarios en los proyectos de Explotación de Información?



## 4. Solución

En este capítulo se desarrolla la solución propuesta para la problemática mencionada en el capítulo anterior realizando una descripción de la solución planteada (sección 4.1) y la síntesis de esta solución en tablas de procesos (sección 4.2).

### 4.1. Propuesta

Se propone como solución a los problemas mencionados en el capítulo 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.

## **4.2. Solución basada en la propuesta**

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 el desarrollo del mismo.

Partiendo del proceso definido para Competisoft, se definió un 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 “Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento”, “Definir ciclos y actividades con base en la descripción del proyecto y en el proceso específico”, “Determinar tiempo estimado para cada actividad”, “Elaborar plan de adquisiciones y capacitación”, “Establecer el calendario de actividades”, “Producir un Plan de Proyecto” y “Formalizar el inicio de un nuevo ciclo del proyecto” del subproceso de “Planificación / Entendimiento del Negocio”.

En el subproceso de “Realización” se definieron las actividades para las tareas “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”, “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” y “Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos”.



De la misma forma se definieron actividades para las tareas “Evaluar el cumplimiento del plan de proyecto y plan de desarrollo” y “Analizar y controlar los riesgos” del subproceso de “Evaluación y control”.

Por último, en el subproceso de “Cierre / Entrega” se definieron actividades para las tareas “Formalizar la terminación del proyecto o ciclo”, “Llevar a cabo el cierre del contrato con subcontratistas” y “Planear la entrega”.

Las actividades propuestas en los párrafos precedentes constituyen otra aportación de esta Tesis, ya que no se encuentran definidas en ninguna de las metodologías evaluadas ni en el modelo de Competisoft.

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.

Estas técnicas incluyen:

- Generación de informes, indicando en cada caso el contenido que deberían incluir los mismos.
- Organigramas y Redes de Expertos en las tareas “Determinar los objetivos del negocio” y “Evaluación de la situación”.
- Análisis de PERT para las tareas “Determinar tiempo estimado para cada actividad” “Establecer el calendario de actividades”.
- Diagramas de Gantt para las tareas “Elaborar plan de adquisiciones y capacitación” y “Establecer el calendario de actividades”.
- DM-COMO o Técnicas empíricas de estimación para la tarea “Calcular el costo estimado del proyecto”.

- Distintos tipos de entrevistas (Estructuradas y No Estructuradas), Taxonomías de Riesgos, Análisis de Riesgo Económico, Análisis de Riesgo Técnico, Análisis de Finanzas, Retorno de la Inversión, Análisis de Riesgo Operativo y de Soporte, Análisis de Riesgo del Programa y uso de Glosarios para la tarea “Evaluación de la situación”.
- Diagramas de Gantt, Análisis de PERT y Análisis de Camino Crítico para las tareas “Producir un Plan de Proyecto” y “Producir un Plan de Desarrollo”.
- Taxonomías de Riesgo para “Analizar y controlar los riesgos”.
- Entrevistas Estructuradas para “Generar el reporte de mediciones y sugerencias de mejora”.

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.

Aunque las tareas son las definidas en dicha metodología y las actividades son las mismas que se describen en CRISP-DM, la organización de estas tareas en un proceso independiente y sus subprocesos es novedosa.

Se agregaron actividades faltantes para obtener un proceso coherente y detallado como puede verse en el subproceso de “Preparación de los datos” cuyas actividades no se mencionan en CRISP-DM, el caso del subproceso de “Evaluación”

en el cual se definió una nueva actividad para obtener los modelos aprobados (lo cual contribuye al mayor grado de formalidad que deseamos imprimir en el nuevo modelo) y la actividad para llevar a cabo la presentación en el subproceso de “Entrega”.

Se incorporaron además 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.

Estas técnicas son novedosas y no se encuentran mencionadas en la metodología CRISP-DM.

Entre estas técnicas se pueden destacar la generación de informes para formalizar distintas tareas junto con una descripción de los datos que estos deben contener, la sugerencia de técnicas como los modelos de Entidad-Relación, Entrevistas Estructuradas y Técnicas Estadísticas para realizar la descripción de los datos, la sugerencia de uso del Método Unificado de Transformación para realizar las transformaciones de los datos o la sugerencia de distintos procesos de Explotación de Información para la construcción de modelos (algoritmos TDIDT, Mapas Auto organizados, Redes Bayesianas o SOM) junto con la especificación de en qué caso debe ser utilizada cada una de estas herramientas.

Como resultado del reordenamiento propuesto y de los aportes de esta Tesis, surgen dos procesos diferenciados. El primero vinculado a la administración de proyectos de Explotación de Información el cual puede verse descrito en la *tabla 4.1*. El segundo de estos procesos relaciona con el desarrollo de proyectos de explotación de información se encuentra resumido en la *tabla 4.2*.

**Tabla 4.1 - Proceso de Administración del Proyecto**

Subproceso: Planificación / Entendimiento del negocio			
Tarea	Actividades	Técnicas	Salida
Determinar los objetivos del negocio	<ol style="list-style-type: none"> <li>1. Desarrollar gráficos organizacionales identificando divisiones, departamentos y grupos de proyecto. Estos gráficos deben identificar los nombres de los directivos y responsables.</li> <li>2. Identificar personas clave en el negocio y sus roles.</li> <li>3. Identificar un sponsor interno (sponsor financiero y usuario principal/experto del dominio).</li> <li>4. Identificar si existe un comité de gobierno y de ser así cuáles son sus miembros.</li> <li>5. Identificar las unidades de negocio que serán impactadas por el proyecto de Explotación de Información.</li> </ol>	<ul style="list-style-type: none"> <li>• Uso de Organigramas y redes de expertos.</li> <li>• Informe indicando personal clave de negocio y roles, sponsors, miembros del comité de gobierno y unidades de negocio impactadas por el proyecto.</li> </ul>	Base de conocimiento del negocio (Background)
	<ol style="list-style-type: none"> <li>1. Describir el problema en términos generales.</li> <li>2. Evaluar el estado actual del proyecto.</li> <li>3. Clarificar los prerequisites del proyecto</li> <li>4. Preparar de ser necesario una presentación sobre que es Explotación de Información para el negocio y presentarla.</li> <li>5. Identificar a los grupos destinatarios del resultado del proyecto (¿se espera un reporte escrito para gerencia o un sistema que va a ser utilizado por usuarios finales?).</li> <li>6. Identificar necesidades y</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo descripción del problema, estado actual del proyecto, prerequisites, requerimientos de negocio y beneficios esperados.</li> <li>• Informe en forma de presentación adecuado a cada grupo identificado de destinatarios del proyecto.</li> </ul>	Objetivos del negocio

	<p>expectativas de los usuarios.</p> <ol style="list-style-type: none"> <li>7. Especificar todas las preguntas de negocio tan precisa y específicamente como sea posible.</li> <li>8. Especificar cualquier otro requerimiento de negocio.</li> <li>9. Especificar los beneficios esperados en término de negocio.</li> </ol>		
	<ol style="list-style-type: none"> <li>1. Especificar los criterios de éxito del negocio.</li> <li>2. Identificar quien determina los criterios de éxito.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe indicando los criterios de éxito del negocio y la persona que los determina.</li> </ul>	Criterios de éxito
Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento	<ol style="list-style-type: none"> <li>1. Evaluar la descripción del proyecto.</li> <li>2. Evaluar el proceso de desarrollo y mantenimiento.</li> <li>3. Definir el proceso específico a utilizar.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe indicando el proceso que se va a utilizar y las justificaciones de su elección.</li> </ul>	Proceso Especifico (forma parte del Plan de Desarrollo)
Definir el protocolo de entrega con el cliente	<ol style="list-style-type: none"> <li>1. Resumir los resultados a desplegar</li> <li>2. Desarrollar y evaluar los planes alternativos de despliegue</li> <li>3. Decidir el plan por cada conocimiento distinto o información resultante <ol style="list-style-type: none"> <li>a. Como va a ser propagada a los usuarios</li> <li>b. Como se va a monitorear el resultado o medir sus beneficios</li> </ol> </li> <li>4. Por cada modelo desplegado o resultado de software: <ol style="list-style-type: none"> <li>a. Como va a ser desplegado dentro de los sistemas de la organización</li> <li>b. Como va a ser monitoreado el uso y medidos sus</li> </ol> </li> </ol>	<ul style="list-style-type: none"> <li>• Informe indicando los resultados de despliegue y planes alternativos, indicar como se propaga la información y como se monitorearan los resultados, indicar posibles problemas.</li> </ul>	Plan de Entrega

	<p style="text-align: center;">beneficios</p> <p>5. Identificar posibles problemas cuando se desplieguen los resultados del proyecto de Explotación de Información</p>		
Definir ciclos y actividades con base en la descripción del proyecto y en el proceso específico	<ol style="list-style-type: none"> <li>1. Evaluar la descripción del proyecto</li> <li>2. Evaluar el proceso específico</li> <li>3. Definir la cantidad de ciclos de desarrollo del proyecto</li> <li>4. Definir actividades para cada uno de los ciclos</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los ciclos definidos junto a las actividades que lo componen.</li> </ul>	Proceso Especifico (forma parte del Plan de Desarrollo)
Determinar tiempo estimado para cada actividad	<ol style="list-style-type: none"> <li>1. Evaluar cada una de las actividades definidas.</li> <li>2. Realizar una estimación de la duración de cada una de las tareas.</li> </ol>	<ul style="list-style-type: none"> <li>• Análisis de PERT</li> </ul>	Calendario de actividades (forma parte del Plan de Desarrollo) incorpora el tiempo estimado en el Plan de Proyecto
Elaborar plan de adquisiciones y capacitación	<ol style="list-style-type: none"> <li>1. Identificar el personal necesario</li> <li>2. Identificar recursos necesarios</li> <li>3. Identificar capacitaciones necesarias para el personal y ciclos en los cuales serán necesarias.</li> <li>4. Establecer ciclos en los cuales son necesarios los recursos</li> <li>5. Elaborar el plan de adquisiciones y capacitación.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe indicando los recursos necesarios y las capacitaciones a realizar.</li> <li>• Diagrama de Gantt par la adquisición de recursos y capacitaciones.</li> </ul>	Plan de Adquisiciones y Capacitación
Establecer el equipo de trabajo	<ol style="list-style-type: none"> <li>1. Establecer la conformación del equipo de trabajo asignando roles y responsabilidades basándose en la descripción del proyecto.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe indicando como se va a componer el equipo de trabajo.</li> </ul>	Equipo de trabajo (forma parte del Plan de Desarrollo)
Establecer el calendario de actividades	<ol style="list-style-type: none"> <li>1. Identificar las actividades necesarias.</li> <li>2. Identificar fecha de inicio y fin de cada una de las actividades.</li> </ol>	<ul style="list-style-type: none"> <li>• Diagrama de Gantt</li> <li>• Análisis de PERT</li> </ul>	Calendario de actividades (forma parte del Plan de Desarrollo)

	<ol style="list-style-type: none"> <li>3. Identificar restricciones en la ejecución de las actividades</li> <li>4. Identificar precedencia en la ejecución de las tareas</li> <li>5. Elaborar el calendario de actividades</li> </ol>		
Calcular el costo estimado del proyecto	<ol style="list-style-type: none"> <li>1. Estimar los costos de recolección de datos.</li> <li>2. Estimar los costos de desarrollar e implementar la solución.</li> <li>3. Identificar beneficios de la solución implementada.</li> <li>4. Estimar costos operativos.</li> </ol>	<ul style="list-style-type: none"> <li>• DM-COMO [Gallego, 2003] / Técnicas empíricas de estimación (Se está trabajando en aproximaciones empíricas que pueden ser utilizadas para la estimación de este tipo de proyectos) [Rodríguez <i>et al.</i>, 2010; Bogado <i>et al.</i>, 2011; Pytel <i>et al.</i>, 2011]</li> <li>• Informe con la estimación de costos y beneficios de la solución</li> </ul>	Costo estimado (forma parte del Plan de Proyecto) / Costos y beneficios
Evaluación de la situación	<ol style="list-style-type: none"> <li>1. Identificar los recursos de hardware disponibles.</li> <li>2. Establecer la disponibilidad de los recursos de hardware para el proyecto.</li> <li>3. Establecer si los planes de mantenimiento de hardware entran en conflicto con la disponibilidad de hardware necesaria para el proyecto de Explotación de Información.</li> <li>4. Identificar la disponibilidad del hardware necesario para las herramientas de Explotación de Información a ser utilizadas (si la herramienta a utilizar ya es conocida en esta etapa).</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo recursos disponibles junto con la necesidad de los mismos y conflictos de disponibilidad. Identificación de fuentes de datos y conocimiento.</li> <li>• Uso de organigramas para la documentación de áreas</li> <li>• Diagramas de expertos para su identificación</li> <li>• Uso de entrevistas estructuradas y no estructuradas para obtener</li> </ul>	Inventario de recursos

	<ol style="list-style-type: none"> <li>5. Identificar las fuentes de datos.</li> <li>6. Identificar los distintos tipos de fuentes de datos.</li> <li>7. Identificar fuentes de conocimiento</li> <li>8. Identificar tipos de fuentes de conocimiento.</li> <li>9. Chequear la disponibilidad de herramientas y técnicas.</li> <li>10. Describir el conocimiento (background) relevante.</li> <li>11. Identificar al sponsor del proyecto.</li> <li>12. Identificar los administradores de sistemas, de bases de datos y personal de soporte técnico para futuras consultas.</li> <li>13. Identificar analistas de mercado, expertos en Explotación de Información y estadistas y chequear sus habilidades.</li> <li>14. Chequear la disponibilidad de los expertos de dominio para fases posteriores.</li> </ol>	información	
	<ol style="list-style-type: none"> <li>1. Especificar las características del grupo al que está destinado.</li> <li>2. Capturar los requerimientos en fechas.</li> <li>3. Capturar los requerimientos de comprensibilidad, precisión, capacidad de deploy, mantenibilidad y repetibilidad del proyecto de Explotación de Información y los modelos resultantes.</li> <li>4. Capturar los requerimientos de seguridad, restricciones legales, privacidad, reporte y fechas del proyecto.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo requerimientos, asunciones y restricciones.</li> </ul>	Requerimientos, suposiciones y restricciones



	<ol style="list-style-type: none"> <li>5. Capturar todas las asunciones (incluidas las implícitas) y hacerlas explícitas.</li> <li>6. Listar las asunciones de calidad de los datos.</li> <li>7. Listar las asunciones de factores externos.</li> <li>8. Clarificar las asunciones que producen cualquiera de las estimaciones.</li> <li>9. Listar todas las asunciones de necesidad de entender y describir o explicar el modelo (ejemplo, como se presenta a la gerencia).</li> <li>10. Chequear las restricciones generales.</li> <li>11. Chequear los derechos de acceso a las fuentes de datos.</li> <li>12. Chequear las técnicas de acceso a los datos (sistema operativo, base de datos, archivos, etc.).</li> <li>13. Chequear si el conocimiento relevante es accesible.</li> <li>14. Chequear las restricciones de presupuesto.</li> </ol>		
	<ol style="list-style-type: none"> <li>1. Identificar riesgos de negocio.</li> <li>2. Identificar riesgos organizacionales.</li> <li>3. Identificar riesgos financieros.</li> <li>4. Identificar riesgos técnicos.</li> <li>5. Identificar riesgos que dependen de los datos y sus fuentes.</li> <li>6. Determinar condiciones bajo las cuales cada riesgo puede ocurrir.</li> <li>7. Desarrollar planes de contingencia.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los riesgos identificados, condiciones de ocurrencia y planes de contingencia.</li> <li>• Taxonomías de riesgos para evaluar</li> <li>• Análisis de Riesgo Técnico</li> <li>• Análisis de Riesgo Económico</li> <li>• Análisis de Finanzas</li> <li>• Retorno de la Inversión.</li> </ul>	<p>Riesgos y contingencias (forma parte del Plan de Proyecto nombrado como Plan de Manejo de Riesgos)</p>

		<ul style="list-style-type: none"> <li>• Análisis de Riesgo Operativo y de Soporte.</li> <li>• Análisis de Riesgo del Programa</li> </ul>	
	<ol style="list-style-type: none"> <li>1. Chequear la disponibilidad de glosarios, si no existieran comenzar un borrador del mismo.</li> <li>2. Hablar con expertos del dominio para entender su terminología.</li> <li>3. Familiarizarse con la terminología de negocio.</li> </ol>	<ul style="list-style-type: none"> <li>• Glosario con la terminología de negocio.</li> </ul>	Terminología
Producir un Plan de Proyecto	<ol style="list-style-type: none"> <li>1. Reunir la información identificada en las tareas anteriores sobre ciclos, actividades, estimaciones de tiempo, planes de adquisición, equipo de trabajo, estimación de costos, calendario, manejo de riesgos y protocolo de entrega.</li> <li>2. Elaborar el plan de proyecto basándose en la información anterior.</li> </ol>	<ul style="list-style-type: none"> <li>• Diagrama de Gantt</li> <li>• Análisis de PERT</li> <li>• Análisis de camino crítico (CPM)</li> <li>• Informe conteniendo la información acumulada.</li> </ul>	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	<ol style="list-style-type: none"> <li>1. Definir el plan de procesos inicial y discutir la factibilidad con todo el personal involucrado.</li> <li>2. Colocar las diferentes metas y técnicas seleccionadas juntas en un procedimiento coherente que solucione las incógnitas del negocio y cumpla con los criterios de éxito del negocio.</li> <li>3. Estimar esfuerzo y recursos necesarios para alcanzar y deployar la solución.</li> <li>4. Identificar los pasos críticos.</li> </ol>	<ul style="list-style-type: none"> <li>• Diagramas de Gantt</li> <li>• Análisis de PERT</li> <li>• Análisis de camino crítico (CPM)</li> <li>• Informe conteniendo la descripción del producto, entregables y equipo de trabajo.</li> </ul>	Plan de Desarrollo (incluye descripción del producto y entregables, proceso específico, equipo de trabajo y calendario)

	<ol style="list-style-type: none"> <li>5. Marcar puntos de decisión.</li> <li>6. Marcar puntos de revisión.</li> <li>7. Identificar las iteraciones mayores</li> </ol>		
	<ol style="list-style-type: none"> <li>1. Crear una lista de criterios de selección para técnicas y herramientas.</li> <li>2. Elegir las técnicas y herramientas potenciales.</li> <li>3. Evaluar que tan apropiadas son las técnicas.</li> <li>4. Revisar y priorizar las técnicas aplicables de acuerdo a la evaluación de soluciones alternativas.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los criterios de selección de técnicas y herramientas con las preferencias de uso.</li> </ul>	Lista inicial de técnicas y herramientas
Formalizar el inicio de un nuevo ciclo del proyecto	<ol style="list-style-type: none"> <li>1. Dar inicio formal al ciclo correspondiente del proyecto.</li> <li>2. Comunicar a los interesados el inicio del nuevo ciclo.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe de inicio del ciclo.</li> </ul>	
<b>Subproceso: Realización</b>			
Tarea	Actividades	Técnicas	Salida
Acordar las tareas con el equipo de trabajo	<ol style="list-style-type: none"> <li>1. Revisar las tareas a realizar en el ciclo actual</li> <li>2. Evaluar los roles necesarios para cada una de las tareas</li> <li>3. Acordar las tareas con el equipo de trabajo</li> <li>4. Distribuir las tareas.</li> </ol>		
Acordar la distribución de información	<ol style="list-style-type: none"> <li>1. Identificar la información requerida para cada tarea</li> <li>2. Distribuir la información al equipo de trabajo según la distribución de tareas.</li> </ol>		
Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario	<ol style="list-style-type: none"> <li>1. Reunirse con el responsable de desarrollo</li> <li>2. Revisar la descripción del producto</li> <li>3. Revisar el equipo de trabajo</li> <li>4. Revisar el calendario</li> <li>5. Realizar correcciones en caso de ser necesario</li> </ol>	<ul style="list-style-type: none"> <li>• Informe incluyendo ajustes necesarios y minuta de reunión.</li> </ul>	
Revisar cumplimiento	<ol style="list-style-type: none"> <li>1. Revisar plan de adquisiciones y</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo</li> </ul>	Reporte de Seguimiento /

del plan de adquisiciones y capacitación	<ol style="list-style-type: none"> <li>capacitación</li> <li>2. Controlar cumplimiento</li> <li>3. Realizar ajustes necesarios.</li> </ol>	información de seguimiento.	Plan de monitoreo y mantenimiento
Administrar subcontratos	<ol style="list-style-type: none"> <li>1. Revisar subcontratos</li> <li>2. Evaluar cumplimiento</li> <li>3. Realizar ajustes necesarios</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo información de seguimiento.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo	<ol style="list-style-type: none"> <li>1. Recolectar reportes de actividades y mediciones</li> <li>2. Recolectar sugerencias de mejoras</li> <li>3. Recolectar productos de trabajo</li> <li>4. Generar un informe con los datos recolectados</li> <li>5. Realizar los ajustes necesarios</li> <li>6. Promover las sugerencias de mejora en caso de ser adecuadas</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los reportes de actividades, mediciones y sugerencias de mejora.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento Reporte de Mediciones y Sugerencias de Mejora
Registrar costo real del proyecto	<ol style="list-style-type: none"> <li>1. Revisar los costos reales del proyecto</li> <li>2. Comparar los costos reales con los estimados</li> <li>3. Registrar los costos</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los costos reales y desviaciones.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Revisar el registro de rastreo basado en los productos de trabajo recolectados	<ol style="list-style-type: none"> <li>1. Revisar los productos de trabajo recolectados</li> <li>2. Revisar el registro de rastreo de los productos</li> <li>3. Realizar ajustes necesarios</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los productos de trabajo recolectados y su registro de seguimiento.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Revisar los productos terminados durante el proyecto	<ol style="list-style-type: none"> <li>1. Revisar los productos terminados durante el proyecto</li> <li>2. Registrar los productos terminados</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los productos terminados.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Recibir y analizar las solicitudes de cambio del cliente	<ol style="list-style-type: none"> <li>1. Recibir solicitudes de cambio del cliente</li> <li>2. Analizar factibilidad de las solicitudes de cambio</li> <li>3. Adecuar el proyecto a las solicitudes de cambio</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo solicitudes de cambio y análisis de las mismas</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Realizar reuniones con el equipo de trabajo y	<ol style="list-style-type: none"> <li>1. Realizar reuniones de avance con el equipo de trabajo</li> <li>2. Realizar reuniones de</li> </ol>	<ul style="list-style-type: none"> <li>• Informes de avance.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y

cliente para reportar avances y tomar acuerdos	avance con el cliente		mantenimiento
<b>Subproceso: Evaluación y control</b>			
Tarea	Actividades	Técnicas	Salida
Evaluar el cumplimiento del plan de proyecto y plan de desarrollo	<ol style="list-style-type: none"> <li>1. Evaluar el cumplimiento del plan de proyecto</li> <li>2. Ajustar el plan de proyecto</li> <li>3. Generar informes de ajuste del plan de proyecto</li> <li>4. Evaluar el cumplimiento del plan de desarrollo</li> <li>5. Ajustar el plan de desarrollo</li> <li>6. Generar informes de ajuste del plan de desarrollo</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo el avance de los planes de proyecto y desarrollo junto con los ajustes realizado y sus justificaciones.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Analizar y controlar los riesgos	<ol style="list-style-type: none"> <li>1. Analizar los riesgos con probabilidad de ocurrencia</li> <li>2. Ejecutar planes de mitigación para los riesgos que sea necesario</li> <li>3. Ejecutar planes de contingencia para los riesgos ocurridos</li> <li>4. Evaluar existencia de nuevos riesgos</li> <li>5. Ajustar el plan de riesgos para los nuevos riesgos</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo información sobre la gestión de riesgos.</li> <li>• Taxonomía de riesgos.</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Generar el reporte de seguimiento del proyecto	<ol style="list-style-type: none"> <li>1. Chequear por aspectos dinámicos (¿qué cosas pueden cambiar en el futuro?)</li> <li>2. Chequear como va a ser monitoreada la precisión</li> <li>3. Evaluar cuando deberían dejar de ser utilizados los resultados del proyecto de Explotación de Información o el modelo, identificar criterios y que debería pasar llegado el momento (actualizar el modelo, ejecutar un nuevo proyecto, etc.)</li> <li>4. Evaluar si los objetivos</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los aspectos dinámicos del proyecto y el plan de monitoreo y mantenimiento</li> </ul>	Reporte de Seguimiento / Plan de monitoreo y mantenimiento

	<p>del negocio y del uso del modelo puede cambiar a lo largo del tiempo. Documentar lo más completamente posible el problema inicial que el modelo está intentando resolver.</p> <p>5. Desarrollar el plan de monitoreo y mantenimiento</p>		
<p>• Subproceso: Cierre / Entrega</p>			
Tarea	Actividades	Técnicas	Salida
Formalizar la terminación del proyecto o ciclo	<ol style="list-style-type: none"> <li>1. Formalizar la terminación del proyecto o ciclo</li> <li>2. Comunicar la finalización a todos los interesados.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe de finalización del proyecto o ciclo.</li> </ul>	Documento de aceptación
Llevar a cabo el cierre del contrato con subcontratistas	<ol style="list-style-type: none"> <li>1. Para cada contrato realizar el cierre del mismo</li> <li>2. Realizar un informe de los contratos</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo la revisión del contrato y aceptación. Incluir un informe de cada subcontratista.</li> </ul>	Informe de Subcontratistas
Generar el reporte de mediciones y sugerencias de mejora	<ol style="list-style-type: none"> <li>1. Entrevistar a todas las personas significativas envueltas en el proyecto y consultarlas acerca de su experiencia a lo largo del proyecto</li> <li>2. Si los usuarios finales en el negocio trabajan con los resultados del proyecto de Explotación de Información, entrevistarlos acerca del nivel de satisfacción, que se podría haber hecho mejor y si necesitan soporte adicional</li> <li>3. Resumir el feedback dado y escribir un documento de la experiencia</li> <li>4. Analizar el proceso (cosas que funcionaron bien, errores cometidos y lecciones aprendidas)</li> <li>5. Documentar el proceso de</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo el resumen de la experiencia.</li> <li>• Entrevistas estructuradas.</li> </ul>	Reporte de mediciones y sugerencia de mejoras - Lecciones Aprendidas / Documentación de la Experiencia

	<p>Explotación de Información específico (como pueden los resultados y la experiencia de aplicar el modelo ser feedback en el proceso)</p> <p>6. Resumir los detalles para hacer la experiencia útil para futuros proyectos</p>		
Planear la entrega	<p>1. Realizar la planificación de la entrega de los productos.</p>	<ul style="list-style-type: none"> <li>Informe conteniendo la información de la planificación de la entrega.</li> </ul>	Plan de entrega (forma parte del Plan de Proyecto nombrado como protocolo de entrega)

**Tabla 4.2** - Proceso de Desarrollo del Proyecto

• Subproceso: Entendimiento del negocio			
Tarea	Actividades	Técnicas	Salida
Determinar las metas del proyecto de Explotación de Información	<p>1. Trasladar las interrogantes del negocio a metas del proyecto de Explotación de Información</p> <p>2. Especificar el o los tipos de problemas de Explotación de Información (descubrimiento de reglas de comportamiento, descubrimiento de grupos, ponderación de interdependencia de atributos, descubrimiento de reglas de pertenencia a grupos o ponderación de reglas de comportamiento).</p>	<ul style="list-style-type: none"> <li>Informe indicando una traducción de los interrogantes a metas de negocio y los tipos de problemas de Explotación de Información</li> </ul>	Metas del proyecto de Explotación de Información
	<p>1. Especificar los criterios de valoración de los modelos.</p> <p>2. Determinar benchmarks para la evaluación de los criterios.</p> <p>3. Especificar criterios que direccionan a valoraciones de criterios</p>	<ul style="list-style-type: none"> <li>Informe conteniendo criterios y benchmarks.</li> </ul>	Criterios de éxito del proyecto de Explotación de Información

	subjetivos (ej. la habilidad de ser explicados).		
<ul style="list-style-type: none"> <li>Subproceso: Entendimiento de los datos</li> </ul>			
Tarea	Actividades	Técnicas	Salida
Reunir los datos iniciales	<ol style="list-style-type: none"> <li>Identificar cual es la información necesaria.</li> <li>Chequear si toda la información necesaria se encuentra actualmente disponible.</li> <li>Especificar el criterio de selección</li> <li>Seleccionar tablas o archivos de interés</li> <li>Seleccionar datos dentro de las tablas o archivos.</li> <li>Pensar en que tanta historia debe ser utilizada (no siempre es necesario utilizar toda la historia disponible).</li> <li>Si los datos contienen entradas de texto libre evaluar si es necesario codificarla para el modelado o su es necesario agrupar entradas específicas.</li> <li>Evaluar cómo se deben adquirir los datos faltantes.</li> <li>Describir como se deben extraer los datos.</li> </ol>	<ul style="list-style-type: none"> <li>Informe conteniendo criterios de selección de información, información seleccionada, límites históricos a utilizar, evaluación de los métodos de obtención de datos faltantes y descripción de los métodos de extracción de los datos.</li> </ul>	Reporte de datos iniciales
Describir los datos	<ol style="list-style-type: none"> <li>Identificar los datos y métodos de captura</li> <li>Acceder a las fuentes de datos</li> <li>Usar análisis estadístico si es apropiado</li> <li>Reportar tablas y sus relaciones</li> <li>Chequear el volumen de los datos, numero de múltiples y complejidad</li> <li>Evaluar si los datos contiene entradas de texto libre</li> <li>Chequear la accesibilidad y disponibilidad de los</li> </ol>	<ul style="list-style-type: none"> <li>Informe conteniendo los datos y métodos de captura, descripción de los tipos de datos, significados de los datos, estadísticas básicas y análisis de los datos.</li> <li>Diagramas Entidad-Relación para documentar tablas y relaciones</li> <li>Entrevistas estructuradas para entrevistar a los</li> </ul>	Reporte de descripción de datos



	<p>atributos</p> <ol style="list-style-type: none"> <li>8. Chequear los tipos de atributos</li> <li>9. Chequear los rangos de valores de los atributos</li> <li>10. Analizar la correlación de atributos</li> <li>11. Entender el significado de cada atributo y valor de atributo en términos del negocio</li> <li>12. Para cada atributo computar estadísticas básicas (por ejemplo: distribución, promedio, máximo, mínimo, etc.)</li> <li>13. Analizar las estadísticas básicas y relacionar los resultados a los significados en términos de negocio.</li> <li>14. Evaluar si el atributo es relevante para la meta específica de Explotación de Información.</li> <li>15. Evaluar si el significado del atributo es usado consistentemente.</li> <li>16. Entrevistar a expertos de dominio en su opinión de la relevancia de los atributos.</li> </ol>	<p>expertos</p> <ul style="list-style-type: none"> <li>• Técnicas estadísticas</li> </ul>	
Explorar los datos	<ol style="list-style-type: none"> <li>1. Analizar propiedades de atributos interesantes en detalle</li> <li>2. Identificar características de sub-poblaciones</li> <li>3. Considerar y evaluar información y resultados en los reportes de descripción de datos</li> <li>4. Formular hipótesis e identificar acciones</li> <li>5. Transformar hipótesis en metas de Explotación de Información si es posible</li> <li>6. Clarificar las metas de Explotación de Información o hacerlas más precisas.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo análisis de las propiedades de los atributos, sub-poblaciones, hipótesis y sus transformaciones a metas.</li> </ul>	Reporte de exploración de datos

	7. Realizar análisis básicos para verificar las hipótesis		
Verificar la calidad de los datos	<ol style="list-style-type: none"> <li>1. Identificar valores especiales y catalogar su significado</li> <li>2. Chequear cubrimiento</li> <li>3. Chequear claves</li> <li>4. Evaluar si el significado de los atributos y los valores encajan entre si</li> <li>5. Identificar atributos faltantes y campos en blanco</li> <li>6. Dar significado a la falta de datos</li> <li>7. Chequear por atributos con valores diferentes que poseen significados similares</li> <li>8. Chequear la escritura de los valores (diferencias en las mayúsculas y minúsculas, uso de sinónimos, et.)</li> <li>9. Chequear desviaciones, decidir cuando una desviación es nociva o puede indicar un fenómeno interesante.</li> <li>10. Chequear por plausibilidad de valores (ejemplo, campos que tienen el mismo valore o muy similar)</li> <li>11. Si los datos están almacenados en archivos planos chequear que delimitadores son usados y si se usan consistentemente en todos los atributos.</li> <li>12. Si lis datos están almacenados en archivos planos chequear el número de campos en cada registro y si coinciden.</li> <li>13. Chequear consistencias y redundancias entre las diferentes fuentes.</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo valores especiales y su significado, identificación de atributos faltantes y campos en blanco junto con su significado, valores diferentes con significado similar, descripción de los tipos de ruido, valores afectados y plan de trabajo con los mismos.</li> </ul>	Reporte de calidad de los datos

	<p>14. Planear como lidiar con ruidos</p> <p>15. Detectar tipos de ruidos y atributos que son afectados por estos.</p>		
<p>• Subproceso: Preparación de los datos</p>			
Tarea	Actividades	Técnicas	Salida
	<p>1. Reunir los datos producidos por la fase de preparación de datos que van a ser utilizados para modelar la mayor parte del trabajo de análisis del proyecto</p>		Datasets
	<p>1. Reunir la descripción de los datos anteriores</p>	<ul style="list-style-type: none"> <li>Informe conteniendo la descripción de los datos.</li> </ul>	Descripción de los Datasets
Seleccionar los datos	<p>2. Recolectar datos adicionales apropiados</p> <p>3. Realizar test de significancia y correlación para decidir si los campos deben ser incluidos</p> <p>4. Reconsiderar los criterios de selección de datos basándose en las experiencias de la calidad de los datos y la exploración de datos</p> <p>5. Reconsiderar los criterios de selección de datos basándose en la experiencia de modelado</p> <p>6. Seleccionar diferentes subconjuntos de datos</p> <p>7. Considerar el uso de técnicas de muestreo</p> <p>8. Documentar las razones para inclusiones o exclusiones</p> <p>9. Chequear las técnicas disponibles para el muestreo de datos</p>	<ul style="list-style-type: none"> <li>Informe conteniendo los criterios de selección reconsiderados, técnicas de muestreo utilizadas y razones de inclusión/exclusión de datos.</li> </ul>	Justificación de inclusión / exclusión
Limpiar los datos	<p>1. Reconsiderar como lidiar con los tipos de ruido observados</p> <p>2. Corregir, remover o ignorar los ruidos</p>	<ul style="list-style-type: none"> <li>Informe conteniendo las correcciones en el trabajo con ruidos y valores especiales.</li> </ul>	Reporte de limpieza de datos

	<ol style="list-style-type: none"> <li>3. Decidir cómo lidiar con valores especiales y su significado.</li> <li>4. Reconsiderar los criterios de selección de datos basándose en la experiencias de limpieza de datos</li> </ol>		
Construir los datos	<ol style="list-style-type: none"> <li>1. Chequear los mecanismos de construcción disponibles con la lista de herramientas sugeridas para el proyecto</li> <li>2. Decidir si es mejor realizar la construcción dentro o fuera de la herramienta</li> <li>3. Reconsiderar los criterios de selección de los datos basándose en las experiencias de construcción de datos</li> </ol>		
	<ol style="list-style-type: none"> <li>1. Decidir si hay atributos que deben ser normalizados</li> <li>2. Considerar el agregado de nueva información sobre la importancia relevante de atributos agregando nuevos atributos</li> <li>3. Decidir cómo pueden ser construidos o imputados los atributos faltantes.</li> <li>4. Agregar nuevos atributos para acceder a los datos</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo las decisiones de normalización, agregado de información y construcción de atributos faltantes.</li> </ul>	Atributos derivados
	<ol style="list-style-type: none"> <li>1. Especificar los pasos de transformaciones necesarias en términos de facilidades de transformación disponibles.</li> <li>2. Realizar los pasos de transformación</li> <li>3. Chequear técnicas disponibles de generación de registros si son necesarias</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los pasos de transformaciones necesarias e identificación de técnicas para generación de registros.</li> <li>• Método Unificado de Transformación (MUT) [Merlino, 2004]</li> </ul>	Registros generados
Integrar los datos	<ol style="list-style-type: none"> <li>1. Chequear si las facilidades de integración</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo modificaciones en los</li> </ul>	Datos combinados

	<p>son capaces de integrar las fuentes de entrada como es requerido</p> <ol style="list-style-type: none"> <li>Integrar las fuentes y almacenar los resultados</li> <li>Reconsiderar los criterios de selección de los datos basándose en las experiencias de integración de datos</li> </ol>	criterios de selección.	(combinación de tablas y agregaciones)
Formatear los datos	<ol style="list-style-type: none"> <li>Reordenar los atributos si es requerido por las herramientas</li> <li>Reordenar los registros si es necesario o conveniente</li> <li>Reformatear valores para satisfacer los requerimientos de la herramienta de modelado</li> <li>Reconsiderar los criterios de selección de datos basándose en la experiencia de limpieza de datos.</li> </ol>	<ul style="list-style-type: none"> <li>Informe conteniendo modificaciones a los criterios de selección.</li> </ul>	Datos formateados
<ul style="list-style-type: none"> <li>Subproceso: Modelado</li> </ul>			
Tarea	Actividades	Técnicas	Salida
Seleccionar la técnica de modelado	<ol style="list-style-type: none"> <li>Decidir sobre la técnica apropiada para ejercitar teniendo en mente la herramienta seleccionada</li> </ol>	<ul style="list-style-type: none"> <li>Informe conteniendo los criterios de selección de la técnica a utilizar.</li> </ul>	Técnica de modelado
	<ol style="list-style-type: none"> <li>Definir cualquier asunción hecha por la técnica acerca de los datos</li> <li>Comparar las asunciones con las hechas en el reporte de descripción de datos</li> <li>Asegurarse que las asunciones se mantienen y regresar a la fase de preparación de datos si es necesario</li> </ol>	<ul style="list-style-type: none"> <li>Informe conteniendo asunciones de la técnica acerca de los datos.</li> </ul>	Suposiciones de modelado
Generar el diseño de test	<ol style="list-style-type: none"> <li>Chequear los diseños de test existentes para cada meta de Explotación de Información en forma separada</li> </ol>	<ul style="list-style-type: none"> <li>Informe conteniendo los pasos para testear cada meta de Explotación de Información.</li> </ul>	Diseño de test

	<ol style="list-style-type: none"> <li>2. Decidir los pasos necesarios</li> <li>3. Preparar los datos requeridos para testear</li> </ol>		
Construir el modelo	<ol style="list-style-type: none"> <li>1. Setear los parámetros iniciales</li> <li>2. Documentar las razones para seleccionar estos valores</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los seteos de parámetros iniciales junto con la justificación de sus valores.</li> </ul>	Establecimiento de parámetros
	<ol style="list-style-type: none"> <li>1. Ejecutar la técnica seleccionada en el conjunto de datos de entrada para producir el modelo</li> <li>2. Post procesar los resultados de Explotación de Información</li> </ol>	<ul style="list-style-type: none"> <li>• Descubrimiento de reglas de comportamiento: Algoritmos TDIDT</li> <li>• Descubrimiento de grupos: Mapas Auto organizados (SOM)</li> <li>• Ponderación de interdependencia de atributos: Redes Bayesianas</li> <li>• Descubrimiento de reglas de pertenencia a grupos: SOM y TDIDT</li> <li>• Ponderación de reglas de comportamiento de la pertenencia a grupos: Redes Bayesianas</li> </ul>	Modelos
	<ol style="list-style-type: none"> <li>1. Describir cualquier característica del modelo actual que pueda ser utilizada en el futuro</li> <li>2. Guardar los seteos de parámetros utilizados para producir el modelo</li> <li>3. Dar una descripción detallada del modelo y cualquier característica especial</li> <li>4. Para modelos basados en reglas, listar las reglas producidas mas cualquier valoración de regla anterior o exactitud de la eficacia general del modelo y su cobertura</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo una descripción del modelo y su comportamiento.</li> </ul>	Descripción del modelo

	<ol style="list-style-type: none"> <li>5. Para modelos opacos listar cualquier información técnica acerca del modelo y cualquier descripción del comportamiento producida por el proceso de modelado</li> <li>6. Describir el comportamiento del modelo y su interpretación</li> <li>7. Guardar conclusiones con respecto a patrones en los datos si existen, a veces el modelo revela hechos importantes acerca de los datos sin separar valoraciones del proceso</li> </ol>		
Evaluar el modelo	<ol style="list-style-type: none"> <li>1. Evaluar los resultados con respecto a los criterios de evaluación</li> <li>2. Testear resultados de acuerdo a la estrategia de testeo</li> <li>3. Comparar la evaluación de resultados y su interpretación</li> <li>4. Crear un ranking de resultados con respecto a los éxitos y criterios de evaluación</li> <li>5. Seleccionar los mejores modelos</li> <li>6. Interpretar los resultados en términos de negocio</li> <li>7. Obtener comentarios de los modelos por expertos del dominio o de los datos</li> <li>8. Chequear la plausibilidad del modelo</li> <li>9. Chequear el impacto para las metas de Explotación de Información</li> <li>10. Chequear el modelo contra la base de conocimiento dada para ver si la información descubierta es novedosa y útil</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo los resultados de la evaluación del modelo.</li> <li>• Entrevistas estructuradas para chequear la información con los expertos.</li> </ul>	Evaluación del modelo

	<ol style="list-style-type: none"> <li>11. Chequear la confiabilidad de los resultados</li> <li>12. Analizar los potenciales de despliegue para cada resultado</li> <li>13. Si existe una descripción verbal del modelo generado valorar las reglas, ¿son lógicas?, ¿son factibles?, ¿son demasiadas o muy pocas? , ¿ofenden el sentido común?</li> <li>14. Valorar los resultados</li> <li>15. Obtener una pauta en como ciertas técnicas de modelado y cierto seteo de parámetros conducen a buenos o malos resultados.</li> </ol>		
	<ol style="list-style-type: none"> <li>1. Ajustar parámetros para obtener un mejor modelo</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo la descripción de los ajustes de parámetros.</li> </ul>	Revisión de los parámetros establecidos
<ul style="list-style-type: none"> <li>• Subproceso: Evaluación</li> </ul>			
Tarea	Actividades	Técnicas	Salida
Evaluar resultados	<ol style="list-style-type: none"> <li>1. Entender los resultados del proyecto de Explotación de Información</li> <li>2. Interpretar los resultados en términos de la aplicación</li> <li>3. Chequear el impacto para las metas de Explotación de Información</li> <li>4. Chequear los resultados del proyecto de Explotación de Información contra la base de conocimiento dada para ver si la información descubierta es novedosa y útil</li> <li>5. Evaluar y valorar los resultados con respecto a los criterios de éxito del negocio</li> <li>6. Comparar la evaluación</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo la evaluación de resultados y conclusiones.</li> </ul>	Evaluación de los resultados de Explotación de Información respecto a los criterios de éxito



	<p>de resultados y su interpretación</p> <ol style="list-style-type: none"> <li>7. Crear un ranking de resultados con respecto a los criterios de éxito del negocio</li> <li>8. Chequear el impacto de los resultados para las metas iniciales de la aplicación</li> <li>9. Evaluar si existen nuevos objetivos de negocio que deban ser evaluados mas tarde en el proyecto o en nuevos proyectos</li> <li>10. Guardar las conclusiones para futuros proyectos de Explotación de Información</li> </ol>		
	<ol style="list-style-type: none"> <li>1. Obtener la aprobación de los modelos</li> </ol>	<ul style="list-style-type: none"> <li>• Informe de aprobación de los modelos.</li> </ul>	Modelos aprobados
Revisar el proceso	<ol style="list-style-type: none"> <li>1. Dar un resumen del proceso de Explotación de Información utilizado</li> <li>2. Analizar el proceso de Explotación de Información, para cada etapa evaluar si fue necesaria y si fue ejecutada en forma optima</li> <li>3. Evaluar como puede ser mejorado</li> <li>4. Identificar fallas</li> <li>5. Identificar pasos faltantes</li> <li>6. Identificar posibles acciones alternativas, caminos inesperados en el proceso</li> <li>7. Revisar los resultados del proyecto de Explotación de Información con respecto a los criterios de éxito del negocio</li> </ol>	<ul style="list-style-type: none"> <li>• Informe de evaluación del proceso y sugerencias de mejora.</li> </ul>	Revisión del proceso
Determinar próximos pasos	<ol style="list-style-type: none"> <li>1. Analizar el potencial de despliegue de cada resultado</li> <li>2. Estimar el potencial de mejora del proceso actual</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo el potencial de mejora y las recomendaciones de</li> </ul>	Lista de posibles decisiones

	<ol style="list-style-type: none"> <li>3. Chequear recursos sobrantes para determinar si permiten realizar iteraciones del proceso adicionales</li> <li>4. Recomendar alternativas de continuidad</li> <li>5. Refinar el plan de proceso</li> </ol>	continuidad.	
	<ol style="list-style-type: none"> <li>1. Ranquear las posibles acciones</li> <li>2. Seleccionar una de las posibles acciones</li> <li>3. Documentar la razón de la elección</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo un ranqueo de las posibles acciones y justificaciones de elección de la seleccionada.</li> </ul>	Decisiones
<ul style="list-style-type: none"> <li>• Subproceso: Entrega</li> </ul>			
Tarea	Actividades	Técnicas	Salida
Producir un reporte final	<ol style="list-style-type: none"> <li>1. Identificar que reportes son necesarios</li> <li>2. Analizar que tan bien fueron alcanzadas las metas del proyecto de Explotación de Información</li> <li>3. Identificar los distintos grupos destinatarios de los reportes</li> <li>4. Esbozar la estructura y contenido de el/los reportes</li> <li>5. Seleccionar los descubrimientos a ser incluidos en el/los reportes</li> <li>6. Escribir el/los reportes</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo el resumen de los logros del proyecto adecuado a cada grupo de destinatario.</li> </ul>	Reporte final
	<ol style="list-style-type: none"> <li>1. Decidir los grupos para la presentación final</li> <li>2. Decidir los ítems del reporte final que van a ser incluidos en la presentación final</li> <li>3. Llevar a cabo la presentación</li> </ol>	<ul style="list-style-type: none"> <li>• Informe conteniendo la presentación final del proyecto.</li> </ul>	Presentación final



## 5. Caso de Estudio

En este capítulo se menciona el caso de estudio piloto para evaluar la viabilidad de la propuesta (sección 5.1). Posteriormente se realiza una comparación entre el uso de la metodología CRISP-DM en un caso real contra el uso del modelo propuesto como solución de la problemática planteada por esta Tesis (sección 5.2). Aquí, se divide entre las diferencias y similitudes observadas en las tres etapas en las que se puede encontrar un proyecto: Preparación e Inicio del Proyecto (sección 5.2.1), Ejecución del Proyecto (sección 5.2.2) y Finalización del Proyecto (sección 5.2.3).

### 5.1. Introducción

En su Tesis de Maestría, Flores [Flores, 2009] 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.

## 5.2. Comparación de uso

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. Las etapas mencionadas son: Inicio, Ejecución y Fin.

Esta comparación puede visualizarse en la *tabla 5.1*, 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.

**Tabla 5.1** - Comparación de uso CRISP-DM vs. Modelo Propuesto

Categoría	Proceso	Subproceso	Flores / CRISP-DM	Modelo de Procesos para Proyectos de Explotación de Información
Alta Dirección	Todos	Todos	No se contempla	Uso de los procesos definidos en Competisoft.
Gerencia	Todos	Todos	No se contempla	Uso de los procesos definidos en 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	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).

			entregas. Plan de entrega.	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.

		Cierre / Entrega	Planear la entrega. Generar el reporte de mediciones y sugerencias de mejora.	Ídem anterior. 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 que incluye tareas propias de la administración	Las tareas de administración se ejecutan en un proceso aparte. Se sigue una guía de actividades para cumplir con las tareas. Posible uso de técnicas sugeridas
		Entendimiento de los datos	Uso del subproceso definido en CRISP-DM	Posible uso de técnicas sugeridas
		Preparación de los datos	Uso del subproceso definido en CRISP-DM	Se sigue una guía de actividades para cumplir con las tareas. Posible uso de técnicas sugeridas
		Modelado	Uso del subproceso definido en CRISP-DM	Posible uso de técnicas sugeridas
		Evaluación	Uso del subproceso definido en CRISP-DM	Se sigue una guía de actividades para cumplir con las tareas. Posible uso de técnicas sugeridas
		Entrega	Uso del subproceso definido en CRISP-DM que incluye tareas propias de la administración	Las tareas de administración se ejecutan en un proceso aparte. Se sigue una guía de actividades para cumplir con las tareas. Posible uso de técnicas sugeridas

### **5.2.1. Preparación e inicio del proyecto**

Dado que el foco de CRISP-DM se encuentra en las etapas de ejecución del proyecto es en esta etapa y en la de finalización en la que se pueden encontrar las diferencias más significativas.

La única tarea propuesta por CRISP-DM previa a la ejecución del proyecto es la tarea de Comprensión del Negocio en la cual, en diferentes actividades, Flores propone los objetivos de su proyecto.

Dentro de los objetivos del negocio plantea las interrogantes del mismo que se buscarán satisfacer con el proyecto, plantea una descripción general del estado actual y lo hecho al momento para solucionar estos interrogantes así como cuales son las soluciones que se están aplicando. Se plantean los objetivos iniciales del proyecto y los criterios de éxito del mismo indicando quienes evaluarán dichos criterios.

En la evaluación de la situación Flores describe el estado actual de las bases de datos y los recursos con los cuales cuenta para llevar a cabo el proyecto junto con los requerimientos, presunciones y restricciones del mismo.

En la definición de los objetivos de la minería de datos plantea todos los objetivos que se buscan cumplir con el proyecto.

Finalmente define un plan de proyecto en el cual se describen las etapas en las que se ira ejecutando el proyecto junto con las herramientas que se utilizarán. Estas etapas son netamente de ejecución y no incluyen elementos de seguimiento y control.



CRISP-DM tiene un único proceso en el cual se realizan todas las tareas propuestas. Por el contrario, en la solución propuesta en esta Tesis, esta tarea es parte de un proceso separado, el de Administración de Proyectos.

Siguiendo el modelo propuesto en esta Tesis se deberán realizar las siguientes tareas:

- Concluida la tarea de Comprensión del Negocio se debe definir un proceso específico basado en la descripción del proyecto, el cual forma parte del plan de desarrollo, denominado en CRISP-DM como plan de proyecto. Al mismo tiempo, se debe definir la cantidad de ciclos en los cuales se desarrolla el proyecto y hacer una evaluación de tiempos necesarios para llevar a cabo cada actividad (posiblemente utilizando análisis de PERT). Esta tarea es utilizada como entrada para la construcción de un calendario de actividades en el cual ya se deben incluir las fechas de inicio y fin de cada una de las mismas identificando la precedencia de las actividades.
- Teniendo una descripción de las actividades que a realizadas se debe elaborar un plan de adquisición de personal y capacitación (uso de diagramas de Gantt) en el cual se distribuye la incorporación del personal afectado al proyecto y su desafectación.
- Conociendo las actividades necesarias para llevar a cabo el proyecto se procede a estimar los costos de desarrollo y operación identificando cuales son los beneficios que conlleva la solución a implementar.
- Las tareas de evaluación de la situación, descripción de técnicas y herramientas, determinación de riesgos y planificación de entrega definidas en CRISP-DM se deben mover, como se indica en el modelo propuesto, al proceso de Administración de Proyecto. Se debe añadir a lo definido la generación de planes de riesgos y contingencias para ser utilizados durante la ejecución del proyecto.

- Se debe generar el plan de proyecto y el plan de desarrollo. CRISP-DM sólo define el plan de proyecto que incluye tareas de los dos procesos (Administración de Proyectos e Implementación de Proyectos). En el plan de proyecto se refleja la información relacionada con cómo se ejecuta el proyecto refiriéndose a la gestión del mismo (ciclos, actividades, tiempos, calendario, incorporación y capacitación de personal, control de costos y de avance, gestión de riesgos y protocolo de entrega). En el plan de desarrollo se incluye la descripción de los productos y entregables a construir, el proceso específico que se va a utilizar, el equipo de trabajo que desarrolla cada actividad y el calendario de actividades. Se definen también los pasos críticos, puntos de decisión y puntos de control junto con el establecimiento de las iteraciones de desarrollo.
- Definir, de manera preliminar, el plan de entrega al cliente.
- Siguiendo el proceso de Administración de Proyectos se debe realizar, en esta etapa, la asignación de roles y responsabilidades del equipo de trabajo.
- El inicio de cada nuevo ciclo del proyecto se formaliza comunicando este hito a todos los interesados.

### **5.2.2. Ejecución del proyecto**

Durante la ejecución del proyecto se deben llevar a cabo diferentes tareas de gestión, control y corrección que no se encuentran descriptas en CRISP-DM pero sí en los modelos de procesos utilizados para el desarrollo de software clásico. Estas tareas permiten la realización de acuerdos, control de desvíos, ajustes de los desvíos, aprovechamiento de oportunidades o control de riesgos entre otras. Todas estas actividades se deben realizar a lo largo de toda la ejecución del proyecto y en paralelo al mismo.

Las tareas descritas en la fase de realización del proyecto no son contempladas específicamente en CRISP-DM, y es por este motivo que no se van a encontrar en la Tesis de Flores.

Siguiendo el modelo de procesos propuesto en esta Tesis se deben realizar las siguientes tareas:

- Obtener documentación que acredite los controles y correcciones ejecutados durante la ejecución del proyecto y los acuerdos formalmente adquiridos.
- Asegurar que cada miembro del equipo conoce sus responsabilidades y las tiene asignadas formalmente lo cual se logra a través de la toma de acuerdos en la tarea de distribución de responsabilidades e información. De la misma forma es necesario asegurar que dichas responsabilidades son comprendidas y aceptadas.
- Acordar con el responsable de desarrollo la descripción del producto, los responsables de cada tarea y el calendario propuesto, generando correcciones siempre que sea necesario.
- Ejecutar controles de adquisiciones y capacitación así como la gestión de subcontratos, si es que los hubiera.
- Realizar controles de los productos de trabajo, en el caso de proyectos de explotación de información se pueden considerar como productos de trabajo a controlar a cada una de las fases de desarrollo del proyecto y al abordaje de cada una de las preguntas que el proyecto debe responder.
- Se deben realizar tareas periódicas de control incluyendo la recolección de reportes de actividades y mediciones y sugerencias de mejora y de productos de trabajo. Estas mediciones no solo permiten controlar el avance del proyecto sino también ajustar el proceso para futuros proyectos. Se deben controlar los costos incurridos periódicamente permitiendo conocer el costo real del proyecto, revisar los registros de productos de trabajo

recolectados y el registro de rastreo de los mismos así como revisar los productos terminados.

- Dentro de las tareas periódicas de administración se incluyen la recepción de solicitudes de cambio del cliente y los ajustes del proyecto necesarios. En un proyecto real los cambios son frecuentes y debe definirse un proceso para la administración del cambio.
- Realizar reuniones de progreso tanto con el equipo de trabajo como con el cliente que generen informes de avance y toma de acuerdos.

En el proceso de administración se encuentra el subproceso de evaluación y control. Este subproceso incluye las tareas de evaluación de los planes de proyecto y desarrollo, la generación de ajustes y desarrollo de informes, el análisis de riesgos, ejecución de planes de mitigación o contingencia, la evaluación de nuevos riesgos y el ajuste del plan de riesgo para incluir estos últimos.

Dentro de este subproceso se puede incluir también la tarea de generación de un reporte de seguimiento que, si bien es mencionada en CRISP-DM en forma reducida, da una idea de los aspectos dinámicos del proyecto que pueden relacionarse con la evaluación de los planes de proyecto y desarrollo. En Flores esta tarea se encuentra enfocada al mantenimiento del proyecto.

En el proceso de desarrollo del proyecto no se encontraran diferencias con el trabajo realizado en la Tesis de Flores (teniendo en cuenta que únicamente se consideró necesario eliminar del proceso de desarrollo del proyecto algunos elementos del inicio y del fin del mismo manteniendo todos los de ejecución).

### **5.2.3. Finalización del proyecto**

En la finalización del proyecto la principal diferencia se encuentra en la solicitud de un cierre formal del proyecto y el cierre formal de subcontratos.

Las tareas de generación de reportes de mediciones y sugerencias de mejoras que son tenidas en cuenta en CRISP-DM se realizaran pero en un proceso separado (el de Administración de Proyectos).

Finalmente en el mismo proceso se debe llevar a cabo la tarea de planificación y ejecución de la entrega de productos al cliente.

## **6. Conclusiones y Aportaciones de esta Tesis**

En este capítulo se mencionan las conclusiones obtenidas a partir del desarrollo de esta Tesis (sección 6.1) comenzando por las conclusiones que justifican la Tesis (sección 6.1.1), realizando a continuación una valoración del problema encontrado (sección 6.1.2), siguiendo con las conclusiones de la construcción de la solución (sección 6.1.3) y continuando con las conclusiones obtenidas en la validación de la propuesta (sección 6.1.4), para finalizar se responden los interrogantes de investigación (sección 6.1.5). Al final del capítulo se mencionan las futuras líneas de investigación a desarrollar a partir de esta Tesis (sección 6.2).

### **6.1. Conclusiones**

A partir de este trabajo de tesis se obtuvieron distintas conclusiones que se exponen a continuación.

#### **6.1.1. Valoración sobre la investigación documental**

En la primera parte de este trabajo se expuso la diferencia entre grandes empresas y pequeñas y medianas empresas que justifican el uso de modelos de procesos personalizados para cada una de ellas. Se obtuvo como consecuencia una justificación de porqué crear un modelo que sea exclusivo de pequeñas y medianas empresas a lo cual se suma la existencia de modelos específicos como Competisoft.

En la segunda parte de este trabajo se realizó una descripción de Competisoft, el cual se encuentra altamente ligado al Modelo de Procesos de Software, un modelo ampliamente utilizado y probado en México, haciendo énfasis en la división existente

entre los distintos niveles de procesos, a saber, de Alta Dirección, Gerencia y Operación. También se remarco la gran diferencia que existe entre los procesos de los proyectos de Explotación de Información y los procesos de los proyectos clásicos de desarrollo de software, de las cuales, la más notoria, es la diferencia de etapas que componen el desarrollo propiamente dicho del producto.

En la tercera parte se presento una descripción de cada una de las tres metodologías más importantes que se utilizan para proyectos de Explotación de Información y de algunas de menor relevancia que, creemos, pueden complementarias, finalizando con una comparación entre las mismas.

En la última parte de este trabajo se realizó una comparación de los elementos existentes en el Modelo de Procesos de Software que pueden ser asociados con los distintos elementos de las tres metodologías identificadas como principales. Esta comparación es la base de la creación de un Modelo de Procesos de Explotación de Información.

En vista de lo desarrollado, es importante la creación de un Modelo de Procesos para Explotación de Información orientado a Pequeñas y Medianas Empresas utilizando como base Competisoft. Esta decisión se apoya en el amplio proceso de investigación desarrollado para la generación de Competisoft y que éste está siendo utilizado por la ISO para la generación de una nueva norma de calidad de proceso orientada a PyMES.

### **6.1.2. Valoración del problema**

A partir de las investigaciones documentales realizadas a lo largo de este trabajo de tesis se detectaron distintos problemas en las herramientas actuales para el

desarrollo de proyectos de Explotación de Información. Estos problemas fueron puntualizados en el Capítulo 3 de este trabajo.

Los problemas mencionados se refieren principalmente a las carencias detectadas en el área de gestión de proyectos y de la empresa en las metodologías utilizadas actualmente para el desarrollo de proyectos de Explotación de Información.

Al mismo tiempo se mencionó la existencia de algunas herramientas básicas de gestión pero integradas siempre en un único proceso de desarrollo, las cuales no alcanzan a cubrir todo el espectro de tareas que deben ser realizadas para mantener un proceso de desarrollo con un alto grado de calidad y previsibilidad.

Otro problema analizado es la no adecuación de los modelos de desarrollo tradicionales a los proyectos de Explotación de Información. Si bien se puede trabajar con estos modelos, el proceso de desarrollo debe llevarse a cabo con alguna de las metodologías existentes, dado que las etapas naturales de los procesos de desarrollo de software clásico son diferentes a las de los proyectos de Explotación de Información.

En el Capítulo 3 de esta Tesis se puntualizaron las diferencias encontradas y se propuso los pasos a seguir para construir un modelo de procesos que solucione estos problemas.

### **6.1.3. Valoración de la solución**

En el Capítulo 4 de este trabajo se propuso una solución para los problemas detectados durante la investigación documental. Esta solución se basó en la



utilización del modelo de procesos Competisoft, utilizado para el desarrollo de proyectos de software clásico, como punto de partida de un nuevo modelo de procesos que se adecúe al desarrollo de proyectos de Explotación de Información.

Al mismo tiempo se utilizó la metodología para desarrollo de proyectos de Explotación de Información CRISP-DM, por ser considerada la más completa de todas las metodologías evaluadas, para aportar las herramientas necesarias para llevar a cabo las adecuaciones necesarias en Competisoft para su uso en proyectos de estas características.

La solución propuesta incluyó el reemplazo del proceso de desarrollo de Competisoft por las fases de desarrollo descritas en CRISP-DM, la propuesta de actividades y herramientas faltantes en estas fases para poder ser consistente con la definición de los procesos de Competisoft.

Al mismo tiempo se incluyeron nuevos elementos en el proceso de administración de Competisoft para soportar las diferencias surgidas por las distintas naturalezas de los proyectos de Explotación de Información y los de software clásico.

En este proceso se incluyeron fases definidas en CRISP-DM y se propusieron actividades y herramientas nuevas que fueron consideradas de utilidad para el modelo.

#### **6.1.4. Valoración del Caso de Estudio**

En el Capítulo 5 de esta Tesis se realizó un caso de estudio realizando una comparación entre los resultados obtenidos en un proyecto de Explotación de Información llevado a cabo utilizando la metodología CRISP-DM y los resultados

que se hubieran obtenido de haber sido realizado con el modelo propuesto como solución en esta Tesis.

La conclusión significativa de esta comparación es la falta de información histórica de la ejecución del proyecto en el caso del desarrollo siguiendo la metodología CRISP-DM lo cual es mejorado en el modelo propuesto a lo largo de todo el proceso de Administración de Proyectos. Esta información permite un control riguroso del estado de avance del proyecto y de la performance del mismo permitiendo que el éxito de un proyecto no sea un caso aislado o basado en la suerte sino un logro de un proceso controlado y mejorado con cada experiencia.

#### **6.1.5. Respuesta a los interrogantes de investigación**

Finalmente, las respuestas a las interrogantes de investigación realizadas en la sección 3.3 pueden ser resumidas en las siguientes conclusiones.

Se logró construir un modelo de procesos para proyectos de Explotación de Información que contempla dos procesos bien diferenciados.

El proceso de Administración de Proyectos (el cual puede verse resumido en la tabla 4.1) cuyo fin es mantener el proceso de desarrollo en movimiento y corregir aquellos desvíos que se produzcan con el fin de lograr una finalización exitosa del proyecto. Además, este proceso, se encarga de recolectar información necesaria para aumentar la calidad del proceso de desarrollo permitiendo realizar ajustes en el mismo y de mantener un estándar en la realización de proyectos convirtiendo un proceso aislado y posiblemente caótico, en un proceso controlado y repetible del cual podemos esperar una finalización exitosa con mayor probabilidad.

El proceso de Desarrollo de Proyectos (el cual puede verse resumido en la tabla 4.1) cuyo fin es la producción del proyecto en sí, e intenta cubrir todas las etapas de ejecución del mismo, desde la toma de requerimientos hasta la entrega del producto final al cliente.

En cuanto a la segunda interrogante de investigación se puede decir que los modelos de procesos creados para desarrollo de proyectos de software clásicos no se ajustan completamente a los proyectos de Explotación de Información. Estos modelos pueden sin embargo abarcar sin problemas muchos de los procesos y tareas referidas a la administración de la empresa (Proceso de Alta Dirección) y de la administración del proyecto.

La conclusión final de este interrogante es que los procesos de los modelos de desarrollo de software clásicos pueden ser adecuados para el trabajo con proyectos de Explotación de Información, sin embargo no pueden ser utilizados completamente sin esta adecuación.

## **6.2. Futuras líneas de trabajo**

En etapas posteriores de trabajo se planea realizar experiencias de trabajo para continuar evaluando la viabilidad de la solución propuesta en esta Tesis.

Dentro de las experiencias futuras a llevar a cabo se prevé la cooperación de los cursos de “Tecnologías de Explotación de Información” de 5to año de Ingeniería en Sistemas de Información de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional y de 5to año de la Licenciatura en Sistemas de la Universidad Nacional de Lanús para formular una primera validación de la viabilidad del modelo propuesto.

De la misma forma se propone evaluar la viabilidad del mismo en proyectos reales una vez constatada la viabilidad en los casos experimentales en el ambiente académico.

## 7. Bibliografía

- Aluja, T. (2001). *La minería de datos, entre la estadística y la Inteligencia Artificial*. *Questiió*, Vol. 25, Nro. 3, Pág. 479-498. <http://upcommons.upc.edu/revistes/bitstream/2099/4162/4/article.pdf>. Página web vigente al 13/09/09.
- Amón, I., Jiménez, C. (2009). *Hacia una Metodología para la Selección de Técnicas de Depuración*. Cuarto Congreso Colombiano de Computación 4CCC, Sociedad Colombiana de Computación, UNAB, UIS. <http://serverlab.unab.edu.co:8080/wikimedia/memorias/fullpapers/15.pdf>. Página web vigente al 26/09/09.
- Azevedo, A., Santos, M. F. (2008). *KDD, SEMMA and CRISP-DM: a parallel overview*. IADIS 2008. [http://www.iadis.net/dl/final\\_uploads/200812P033.pdf](http://www.iadis.net/dl/final_uploads/200812P033.pdf). Página web vigente al 5/09/09.
- Bogado, V., Dapozo, G., García-Martínez, R. (2011). *Estimación en Gestión de Proyectos de Software Basada en Explotación de Información*. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Pág. 616-620. ISBN 978-950-673-892-1. <http://iidia.com.ar/rgm/comunicaciones/WICC-2011-3894.pdf>. Página web vigente al 24/07/11.
- Britos, P. (2005). *Objetivos del Negocio y Procesos de Minería de Datos Basados en Sistemas Inteligentes*. Reportes Técnicos de Ingeniería de Software. Vol. 7. Nro. 1. Pág. 26 – 29. <http://laboratorios.fi.uba.ar/lsi/R-rtis-7-1-2005-26-29.pdf>. Página web vigente al 19/09/09.
- Britos, P. (2008). *Procesos de Explotación de Información basados en Sistemas Inteligentes*. Tesis Doctoral. Universidad Nacional de La Plata,

---

Facultad de Informática. La Plata, Argentina. <http://www.iidia.com.ar/rgm/tesistas/td-pb-fi-unlp.pdf>. Página web vigente al 5/09/09.

Britos, P., Dieste, O., García-Martínez, R. (2008). *Requeriments Elicitation in Data Mining for Business Intelligence Projects*. Advances in Information Systems Research, Education and Practice. David Avison, George M. Kasper, Barbara Pernici, Isabel Ramos, Dewald Roode Eds. (Boston: Springer), IFIP Series, 274: 139-150.

Carnegie Mellon University, Software Engineering Institute (SEI) (2006). *CMMI-DEV for Development, Version 1.2*. <http://www.sei.cmu.edu/library/abstracts/reports/06tr008.cfm>. Página web vigente al 5/09/09.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step Data Mining guide*. U.S.A. <http://www.crisp-dm.org/>. Página web vigente al 1/09/09.

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. Dirección de Postgrado. <http://www.iidia.com.ar/rgm/tesistas/flores-tesisdemagister.pdf>. Página web vigente al 24/07/11.

Gallego, O. (2003). *Modelo Matemático Paramétrico de Estimación para Proyectos de Data Mining (DMCOMO)*. Tesis Doctoral. Universidad Politécnica de Madrid. Facultad de Informática. Departamento de Lenguajes y Sistemas e Ingeniería de Software. <http://oa.upm.es/282/1/10200305.pdf>. Página web vigente al 24/07/11.

García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cataneo, F., Rodriguez, D., Pytel, P., Vanrell, J. A. (2011). *Towards and*

---

*Information Mining Engineering*. Software Engineering, Methods, Modeling and Teaching. Pág. 83-99. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6.

Jackson, J. (2002). *Data Mining: A Conceptual Overview*. Communications of the Association for Information Systems. Vol. 8. Pág. 267-296. [http://faculty.wiu.edu/C-Amaravadi/is524/res/dm\\_c\\_ov.pdf](http://faculty.wiu.edu/C-Amaravadi/is524/res/dm_c_ov.pdf). Página web vigente al 13/09/09.

Kuna, H., García-Martínez, R., Villatoro Machuca, F. (2009). *Procedimientos de la Explotación de Información para la Identificación de Datos Faltantes, con Ruido e Inconsistentes*. Proceedings XI Workshop de Investigadores en Ciencias de la Computación. Pág. 236-238. <http://laboratorios.fi.uba.ar/lisi/rgm/comunicaciones/WICC-2009-236-238.pdf>. Página web vigente al 13/09/09.

Larose, D. T. (2005). *Discovering Knowledge in Data, an introduction to Data Mining*. John Wiley & Sons. Estados Unidos. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.589&rep=rep1&type=pdf>. Página web vigente al 28/08/09.

Mas, A. y Amengual, E. (2005). *La mejora de los procesos de software en las pequeñas y medianas empresas (PYME). Un nuevo modelo y su aplicación a un caso real*. REICIS, Revista Española de Innovación, Calidad e Ingeniería del Software, Diciembre, Vol. 1, Nro. 2, Pág. 7-29. <http://redalyc.uaemex.mx/redalyc/pdf/922/92210203.pdf>. Página web vigente al 3/09/09.

Mendez, P. D., Rodríguez, A. D. (2009). *Herramienta de Estudio de Viabilidad para Proyectos que Utilizan la Metodología P<sup>3</sup>TQ*. Trabajo Profesional de Ingeniería en Informática. Laboratorio de Sistemas Inteligentes.

Facultad de Ingeniería. Universidad de Buenos Aires.  
<http://laboratorios.fi.uba.ar/lsi/mendez-rodriguez-trabajoprofesional.pdf>. Página web vigente al 1/09/09.

Merlino, H. (2004). *Un Método de Procesamiento de Datos Orientado al Uso de Explotación de Información Basado en Sistemas Inteligentes*. Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos. Instituto Tecnológico de Buenos Aires. <http://www.itba.edu.ar/nuevo/archivos/secciones/merlino-trabajofinaldeespecialidad.pdf>. Página web vigente al 24/07/11.

Ochoa, A. (2006). *Uso de Técnicas de Educación para el Entendimiento de Negocio*. Tesis de Maestría. Universidad Politécnica de Madrid. <http://www.iidia.com.ar/rgm/tesistas/ochoa-tesisdemagister.pdf>. Página web vigente al 20/09/09.

Ochoa, A. (2005). *Propuesta de Técnicas de Protofase Metodológica para la Comprensión del Negocio*. Reportes Técnicos en Ingeniería del Software, Vol. 7, Nro. 1. Pág. 21 - 25. <http://laboratorios.fi.uba.ar/lsi/R-rtis-7-1-2005-21-25.pdf>. Página web vigente al 19/09/09.

Oktaba, H., Piattini, M., Pino, F.J., Orozco, M.J. y Alquicira, C. (2008). *Competisoft, Mejora de Procesos Software para Pequeñas y Medianas Empresas y Proyectos*. Ra-Ma.

Oktaba, H., García, F., Piattini, M., Ruiz, F., Pino y F.J., Alquicira, C. (2007). *Software Process Improvement: The Competisoft Project*. Computer Vol. 40, Nro. 10, Pág. 21 - 28. Madrid, España. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4343684](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4343684).  
Página web vigente al 3/09/09.



- Oktaba, H., Alquicira Esquivel, C., Ramos, A. S., Martínez Martínez, A., Quintanilla Ozorio, G., Ruvalcaba López, M., López Lira Hinojo, F., Rivera López, M. E., Orozco Mendoza, M. J., Fernández Ordoñez, Y. y Flores Lemus, M. A. (2005). *Modelo de Procesos para la Industria de Software*. Secretaría de Economía de México. <http://www.comunidadmoprosoft.org.mx/>. Página web vigente al 1/09/09.
- Oktaba, H., Alquicira Esquivel, C., Ramos, A. S., Palacios Elizalde, J., Pérez Escobar, C. J. y López Lira Hinojo, F. (2004). *Método de Evaluación de Procesos para la Industria de Software*. Secretaría de Economía de México. <http://www.software.net.mx/NR/rdonlyres/ED7B3399-0CA4-412E-9FAC-0EEB94F85C5F/1224/EvalProSoftv11.pdf>.  
Página web vigente al 1/09/09.
- Pino, F. J., García F. y Piattini, M. (2006). *Revisión sistemática de mejora de procesos software en micro, pequeñas y medianas empresas*. Revista Española de Innovación, Calidad e Ingeniería de Software, Vol. 2, Nro.1. <http://redalyc.uaemex.mx/redalyc/pdf/922/92220103.pdf>.  
Página web vigente al 3/09/09.
- Pytel, P., Tomasello, M., Rrodríguez, D., Arboleya, H., Pollo-Cattaneo, M., Britos, P., García-Martínez, R. (2011). *Estimación de Proyectos de Explotación de Información. Estudio Comparado de Modelos Analíticos y Empíricos*. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Pág. 295-299. ISBN 978-950-673-892-1. <http://idia.com.ar/rgm/comunicaciones/WICC-2011-3848.pdf>. Página web vigente al 24/07/11.
- Rabuñal Dopico, J. R., (2002). *Metodología para el Desarrollo de Sistemas de Extracción de Conocimiento en RNA*. Tesis Doctoral. Universidad da

Coruña, Facultad de Informática, Departamento de Tecnoloxías da Información e as Comunicaci3ns. <http://sabia.tic.udc.es/articulos/2002/Tesis%20Juanra.pdf>. P3gina web vigente al 27/09/09.

Rodr3guez, D., Pollo-Cattaneo, F., Britos, P., Garc3a-Mart3nez, R. (2010). *Estimaci3n Emp3rica de Carga de Trabajo en Proyectos de Explotaci3n de Informaci3n*. Proceedings XIII Workshop de Investigadores en Ciencias de la Computaci3n. P3g. 616-620. ISBN 978-950-673-892-1. <http://iidia.com.ar/rgm/comunicaciones/WICC-2011-3894.pdf>. P3gina web vigente al 24/07/11.

SAS Institute Inc. (1998). *Data Mining and the Case for Sampling Solving Business Problems Using SAS® Enterprise Miner™ Software*. <http://www.sasenterpriseminer.com/documents/SAS-SEMMA.pdf>. P3gina web vigente al 12/09/09.

SEI (2006). *CMMI® for Development, Version 1.2*. Carnegie Mellon University, Software Engineering Institute. <http://www.sei.cmu.edu/reports/06tr008.pdf>. P3gina web vigente al 24/07/11.

Sumathi, S. y Sivanandam, S. N. (2006). *Introduction to Data Mining and its Applications*. Studies in Computational Intelligence (SCI). Vol. 29, P3g. 387-409. Springer Berlin / Heidelberg <http://www.springerlink.com/content/j128n719h46u5455/?p=23d78a05299845cf96fd5a35c8a0d653&pi=7>. P3gina web vigente al 01/09/09.

Vanrell, J. A., Bertone, R., Garc3a-Mart3nez, R. (2010a). *Modelo de Proceso de Operaci3n para Proyectos de Explotaci3n de Informaci3n*. Anales del XVI Congreso Argentino de Ciencias de la Computaci3n. P3g. 674-682. ISBN 978-950-9474-49-9.

---

Vanrell, J. A., García-Martínez, R., Bertone, R. (2010b). *Un Modelo de Procesos de Explotación de Información*. Proceedings XII Workshop de Investigadores en Ciencias de la Computación. Pág. 167-171. ISBN 978-950-34-0652-6.

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. <http://posgrado.frba.utn.edu.ar/investigacion/especialidades/vanrell-trabajofinaldeespecialidad.pdf>. Página web vigente al 24/07/10.