



Repositorio Digital Institucional
"José María Rosa"

Universidad Nacional de Lanús
Secretaría Académica
Dirección de Biblioteca y Servicios de Información Documental

Hernan Leandro Arboleya

Propuesta de ciclo de vida y mapa de actividades para proyectos de explotación de información

Trabajo Final Integrador presentado para la obtención del título de Licenciado en Sistemas

Director de Trabajo Final Integrador

Ramón García Martínez

El presente documento integra el Repositorio Digital Institucional "José María Rosa" de la Biblioteca "Rodolfo Puiggrós" de la Universidad Nacional de Lanús (UNLa)

This document is part of the Institutional Digital Repository "José María Rosa" of the Library "Rodolfo Puiggrós" of the University National of Lanús (UNLa)

Cita sugerida

Arboleya, Hernan Leandro. (2013). Propuesta de ciclo de vida y mapa de actividades para proyectos de explotación de información [en Línea]. Universidad Nacional de Lanús. Departamento de Desarrollo Productivo y Tecnológico

Disponible en: http://www.repositoriojmr.unla.edu.ar/download/TFI/LicSis/033876_Arboleya.pdf

Condiciones de uso

www.repositoriojmr.unla.edu.ar/condicionesdeuso



www.unla.edu.ar
www.repositoriojmr.unla.edu.ar
repositoriojmr@unla.edu.ar



Propuesta de Ciclo de Vida y Mapa de Actividades para Proyectos de Explotación de Información

Alumno

APU Hernán ARBOLEYA

Directores

Dr. Ramon Garcia Martinez y Mg. Dario Rodriguez

TRABAJO FINAL PRESENTADO PARA OBTENER EL GRADO
DE
LICENCIADO EN SISTEMAS

**DEPARTAMENTO
DE DESARROLLO PRODUCTIVO Y TECNOLOGICO
UNIVERSIDAD NACIONAL DE LANUS**

2013

AGRADECIMIENTOS

A mi familia, mis amigos y mis compañeros
por darme apoyo en todos los aspectos de mi vida, tanto personal como profesional.

A mis directores de trabajo final
por la ayuda brindada durante el transcurso de la carrera.

RESUMEN

La Explotación de la Información es una sub-disciplina informática que le aporta herramientas a la Inteligencia del Negocio para poder transformar la información en conocimiento que sirva de base para llevar a cabo la toma de decisiones dentro de las organizaciones. Consiste en la extracción de conocimiento no-trivial que reside de manera implícita en los datos disponibles en distintas fuentes de información. Dicho conocimiento es previamente desconocido y puede resultar útil en procesos de decisión de las organizaciones. Dado que este tipo de proyectos se diferencian a los proyectos convencionales de construcción de software y que la metodologías existente para Proyectos de Explotación de Información no contemplan ni tiene en cuenta los riesgos que dentro de los mismos pueden ocurrir, en este trabajo se propone el diseño de un Modelo de Ciclo de Vida específico para este tipo de proyectos que tenga un análisis de los riesgos. Posteriormente se plantea la construcción de un mapa de actividades teniendo en cuenta el Modelo de Ciclo de Vida desarrollado y los procesos del Modelo de Procesos seleccionado.

ABSTRACT

The Information Mining is a sub-discipline of Corpus of Information Systems that provides tools for Business Intelligence to transform information into knowledge as a basis for carrying out decision making within organizations. Involves the obtention of non-trivial knowledge that resides implicitly in the data available on different sources of information. Such knowledge is previously unknown and may be useful in decision-making processes of organizations. Since Information Mining projects are different to conventional ones and that existing methodologies for software project development do not include or take into account the risks within them can occur, this paper proposes the design of a Life Cycle Model specific for this type of projects that have risks into account. Also is proposed an activities map that uses the proposed life cycle and the Vanrell Process Model.

ÍNDICE

1. INTRODUCCIÓN	1
1.1. Objetivos	1
1.1.1. Objetivos Generales	1
1.1.2. Objetivos Específicos	1
1.2. Alcance	2
1.3. Fundamento del Trabajo	2
1.4. Metodología Empleada	3
1.5. Estructura de este Trabajo	4
2. ESTADO DEL ARTE	7
2.1. Introducción	7
2.2. Modelos de Ciclo de Vida de Interés para el Trabajo Final	8
2.2.1. Modelo de Ciclo de Vida en Espiral de Boehm	9
2.2.2. CRISP-DM	11
2.3. Modelo de Procesos de Interés para el Trabajo Final.	18
2.3.1. Competisoft	18
2.3.1.1. Alta Dirección	19
2.3.1.2. Gerencia	20
2.3.1.3. Operación	20
2.3.2. Modelo de Procesos	21
3. PROBLEMA	29
3.1. Contexto del Problema	29
3.2. Descripción del Problema	29
3.3. Preguntas de Investigación	32
4. SOLUCIÓN	33
4.1. Propuesta	33
4.2. Solución basada en la propuesta	33
4.2.1. Modelo de Ciclo de Vida Propuesto	34
4.2.1.1. Primera Vuelta	35
4.2.1.2. Segunda Vuelta	36
4.2.1.3. Tercera Vuelta	38
4.2.1.4. Cuarta Vuelta	40
4.2.1.5. Quinta Vuelta e iteraciones	41

4.2.1.6. Ultima Vuelta	43
4.2.1.7. División por Fase CRISP-DM	44
4.2.2. Mapa de Actividades Propuesto	45
4.2.2.1. Mapa de Actividades Completo	46
4.2.2.2. Mapa de Actividades – Primera Vuelta del Espiral	46
4.2.2.3. Mapa de Actividades – Segunda Vuelta del Espiral	50
4.2.2.4. Mapa de Actividades – Tercera Vuelta del Espiral	55
4.2.2.5. Mapa de Actividades – Cuarta Vuelta del Espiral	59
4.2.2.6. Mapa de Actividades – Quinta Vuelta del Espiral	61
4.2.2.7. Mapa de Actividades – Ultima Vuelta del Espiral	65
5. PRUEBAS DE CONCEPTO	71
5.1. Presentación de Proyecto para Prueba de Concepto	71
5.2. Elección de Modelo de Ciclo de Vida para el Proyecto	71
5.3. Desarrollo del Proyecto de Explotación de Información	72
5.3.1. Procesos de Administración de Proyectos	72
5.3.1.1. Entendimiento del Negocio	72
5.3.1.1.1. Determinar los Objetivos del Negocio	73
5.3.1.1.1.1. Base de Conocimiento	73
5.3.1.1.1.2. Objetivos del Negocio	75
5.3.1.1.1.3. Criterios del Éxito	77
5.3.1.1.2. Definir el Proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento	77
5.3.1.1.3. Definir ciclos y actividades con base en la descripción del proyecto y en el proceso específico	78
5.3.1.1.4. Determinar tiempo estimado para cada actividad	79
5.3.1.1.5. Elaborar Plan de Adquisiciones y capacitación	79
5.3.1.1.6. Establecer el equipo de trabajo	80
5.3.1.1.7. Establecer el calendario de actividades	80
5.3.1.1.8. Calcular el costo estimado del Proyecto	80
5.3.1.1.9. Evaluación de la Situación	80
5.3.1.1.10. Producir un Plan de Proyecto	82
5.3.1.1.11. Formalizar el inicio de un nuevo ciclo del Proyecto	84
5.3.1.2. Realización	84
5.3.1.2.1. Acordar las tareas con el equipo de trabajo	85

5.3.1.2.2. Acordar la distribución de la información	85
5.3.1.2.3. Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario	85
5.3.1.2.4. Revisar el cumplimiento del plan de adquisiciones y capacitación	85
5.3.1.2.5. Recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo	86
5.3.1.2.6. Registrar costo real del Proyecto	86
5.3.1.2.7. Revisar el registro de rastreo basado en los productos de trabajo recolectados	87
5.3.1.2.8. Registrar los productos terminados durante el Proyecto	87
5.3.1.2.9. Recibir y analizar las solicitudes de cambio del cliente	87
5.3.1.2.10. Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos	87
5.3.1.3. Evaluación y Control	87
5.3.1.3.1. Evaluar el cumplimiento del plan de proyecto y plan de desarrollo	87
5.3.1.3.2. Analizar y controlar riesgos	88
5.3.1.3.3. Generar el reporte de seguimiento del Proyecto	89
5.3.1.4. Cierre/Entrega	89
5.3.1.4.1. Formalizar la terminación del proyecto o ciclo	89
5.3.1.4.2. Llevar a cabo el cierre del contrato con subcontratistas	90
5.3.1.4.3. Generar el reporte de mediciones y sugerencias de mejora	91
5.3.1.4.4. Planear la entrega	92
5.3.2. Procesos de Desarrollo de Proyectos	92
5.3.2.1. Entendimiento del Negocio	92
5.3.2.1.1. Determinar las metas del proyecto de Explotación de Información	92
5.3.2.2. Entendimiento de los Datos	93
5.3.2.2.1. Reunir los datos iniciales	94
5.3.2.2.2. Describir los datos	95
5.3.2.2.3. Explorar los datos	96
5.3.2.2.4. Verificar la calidad de los datos	97
5.3.2.3. Preparación de los Datos	98
5.3.2.3.1. Tareas preparatorias	99
5.3.2.3.2. Seleccionar los datos	99
5.3.2.3.3. Limpiar los datos	100
5.3.2.3.4. Construir los datos	100

5.3.2.3.5. Integrar los datos	101
5.3.2.3.6. Formatear los datos	101
5.3.2.4. Modelado	102
5.3.2.4.1. Seleccionar la técnica de modelado	102
5.3.2.4.2. Generar el diseño del test	103
5.3.2.4.3. Construir el modelo	104
5.3.2.4.4. Evaluar el modelo	107
5.3.2.5. Evaluación	109
5.3.2.5.1. Evaluar resultados	109
5.3.2.5.2. Revisar el proceso	110
5.3.2.5.3. Determinar próximos pasos	112
5.3.2.6. Entrega	112
5.3.2.6.1. Producir un reporte final	112
6. CONCLUSIONES	115
6.1. Aportes del Trabajo Final de Licenciatura	115
6.2. Futuras líneas de Trabajo	116
7. REFERENCIAS	119

ÍNDICE DE FIGURAS

Figura 2.1	Modelo de Ciclo de Vida en Espiral de Boehm	10
Figura 2.2	Metodología CRISP-DM	13
Figura 2.3	Niveles de abstracción de CRISP-DM	14
Figura 4.1	Modelo de Ciclo de Vida Propuesto	37
Figura 4.2	Separación del Modelo de Ciclo de Vida por procesos del Modelo de Procesos	45
Figura 4.3.a)	Mapa de Actividades Propuesto para Administración de Proyectos	47
Figura 4.3.b)	Mapa de Actividades Propuesto para Desarrollo de Proyectos	48
Figura 4.4	Mapa de Actividades Propuesto para la primera vuelta del espiral	51
Figura 4.5	Mapa de Actividades Propuesto para la segunda vuelta del espiral	54
Figura 4.6	Mapa de Actividades Propuesto para la tercera vuelta del espiral	58
Figura 4.7	Mapa de Actividades Propuesto para la cuarta vuelta del espiral	62
Figura 4.8	Mapa de Actividades Propuesto para la quinta vuelta del espiral	64
Figura 4.9	Mapa de Actividades Propuesto para la última vuelta del espiral	68
Figura 5.1	Organigrama de la Organización	74
Figura 5.2	Organigrama del Ministerio de Salud	75
Figura 5.3	GANTT de los ciclos	79
Figura 5.4	GANTT de los ciclos	83
Figura 5.5	Atributos dentro del Input	104
	Marco Contextual Base – Cajeros Automáticos Conectados a EBX”	
Figura 5.6	Target	105
Figura 5.7	Árbol de Decisión	105
Figura 5.8	Información del Modelo	106
Figura 5.9	Parámetros ajustados	108
Figura 5.10	Modelo con parámetros ajustados	111

ÍNDICE DE TABLAS

Tabla 2.1	Fase Entendimiento del Negocio.....	15
Tabla 2.2	Fase Entendimiento de los Datos.....	15
Tabla 2.3	Fase Preparación de los Datos.....	16
Tabla 2.4	Fase Modelado.....	16
Tabla 2.5	Fase Evaluación.....	16
Tabla 2.6	Fase Despliegue.....	17
Tabla 2.7.a)	Procesos de Administración.....	25
Tabla 2.7.b)	Procesos de Administración.....	26
Tabla 2.8	Procesos de Desarrollo.....	26
Tabla 3.1	Relación entre Problemas/riesgos identificados y conceptos a ser educados.....	30
Tabla 4.1	Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR1.....	36
Tabla 4.2	Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR2.....	38
Tabla 4.3	Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR3.....	39
Tabla 4.4	Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR4.....	41
Tabla 4.5	Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR5.....	42
Tabla 4.6	Descripción de los Códigos de las Fases utilizados en la figura 4.3.....	48
Tabla 5.1	Plan de Contingencia.....	81
Tabla 5.2	Plan de Contingencia.....	84
Tabla 5.3	Plan de Contingencia.....	88
Tabla 5.4	Plan de Contingencia para Nuevo Riesgo.....	89
Tabla 5.5	Plan de Contingencia Final.....	90
Tabla 5.6	Análisis de Riesgo para Entendimiento del Negocio.....	94
Tabla 5.7	Definición de Datos.....	97
Tabla 5.8	Posible valores para Datos.....	98
Tabla 5.9	Análisis de Riesgo para Entendimiento de los Datos.....	98
Tabla 5.10	Posibles valores para Datos.....	99
Tabla 5.11	Posible valores para Datos.....	102
Tabla 5.12	Análisis de Riesgo para Preparación de Datos.....	102
Tabla 5.13	Análisis de Riesgo para Modelado.....	109

NOMENCLATURA

ACST	Análisis Cognitivo de los Segmentos de Texto
CompetiSoft	Modelo de Procesos para el Desarrollo de Software basado en MoProSoft
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
E.I.	Explotación de Información
Gantt	Diagramas 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
Null	Sin valor
OPE	Categoría de Competisoft para Operaciones
TDIDT	Top Down Induction of Decision Trees. Inducción de Arboles de Decisión.

1. Introducción

En este capítulo se presenta una introducción al trabajo que se ha realizado. En primer lugar se mencionan objetivos (sección 1.1.) tanto generales (subseccion 1.1.1.) como específicos (subseccion 1.1.2.). En segundo lugar se define el alcance que va a tener (sección 1.2.), se evalúan los fundamentos que dan soporte al mismo (sección 1.3.) y se describe la metodología empleada para poder llevar a cabo el presente trabajo (sección 1.4.). El capítulo finaliza con una descripción de la estructura que presenta el Trabajo.

1.1. Objetivos

En esta sección se lleva a cabo la presentación, en primer lugar, de los objetivos generales que tiene el presente trabajo (subseccion 1.1.1.) y, posteriormente, se presenta una lista de objetivos específicos (subseccion 1.1.2.)

1.1.1. Objetivos Generales

Este trabajo tiene como objetivo general la construcción de un Modelo de Ciclo de Vida que contemple los riesgos existentes en cualquier Proyecto de Explotación de Información, teniendo como base la metodología CRISP-DM [Chapman *et al.*, 2000]. y, por sobre todo, las mejoras que le ha hecho Vanrell a dicha metodología con la presentación de su Modelo de Procesos [Vanrell, 2012]. Adicionalmente, se plantea como objetivo general lograr un Mapa de Actividades teniendo como ejes el Modelo de Ciclo de Vida propuesto y el Modelo de Procesos seleccionado en este trabajo [Vanrell, 2012].

1.1.2. Objetivos Específicos

En el presente trabajo se han identificado una serie de objetivos específicos a partir de los objetivos generales descritos en la subseccion anterior. Los mismos se mencionan a continuación:

- a) Poder describir la metodología CRISP-DM [Chapman *et al.*, 2000] existente para Proyectos de Explotación de Información, mencionar sus alcances, ventajas y sus fases;
- b) Intentar describir a los Modelos de Ciclo de Vida existentes para desarrollo de software convencional, para identificar ventajas y desventajas de cada uno y poder seleccionar cual servirá de base para el diseño del modelo propio para Proyectos de Explotación de Información;
- c) Identificar y describir los procesos del Modelo de Procesos seleccionado;

- d) Seleccionar las tareas adecuadas de la metodología CRISP-DM [Chapman *et al.*, 2000] y de las fases convencionales de cualquier sistema software tradicional para la construcción del Modelo de Ciclo de Vida;
- e) Construir el Modelo de Ciclo de Vida que tenga en cuenta los riesgos y sirva como base para el diseño de estos particulares estilos de proyectos;
- f) Diseñar a partir del Modelo desarrollado y del Modelo de Proceso seleccionado y descrito un Mapa de Actividades donde se marcan que actividades del proceso software se va a ejecutar en un determinado tiempo; y
- g) Llevar a cabo una evaluación de la viabilidad de la propuesta.

1.2. Alcance

El alcance de este Trabajo Final de Licenciatura es establecer las falencias que posee CRISP-DM para la realización de Proyectos de Explotación de Información justificando la creación de un Modelo de Ciclo de Vida que contemple los riesgos que dentro del Proyecto pueden llegar a ocurrir. Adicionalmente se lleva a cabo la descripción de los Modelos de Ciclo de Vida existentes tanto para Proyectos Convencionales de Desarrollo de Software como así también para Proyectos de Explotación de Información.

Se incluye, posteriormente, una descripción y análisis de los Modelos de Procesos de interés para este Trabajo Final.

Al mismo tiempo se incluye la construcción de un Mapa de Actividades que tenga como ejes el Modelo de Ciclo de Vida propuesto y el Modelo de Procesos seleccionado dentro del Trabajo.

Finalmente se evalúa la viabilidad de lo desarrollado en este Trabajo aplicando el Modelo de Ciclo de Vida y el Mapa de Actividades propuesto a 2 casos de estudio.

1.3. Fundamento del Trabajo

En las empresas del mundo se desarrollan distintos tipos de proyectos. Dentro de estos 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. Estos proyectos son especiales, poseen características únicas que lo diferencian de los proyectos comunes de sistemas.

Para los desarrollos de este estilo de proyectos se cuentan actualmente con distintos tipos de técnicas, metodologías y herramientas que soportan la puesta en marcha de los mismos (como por ejemplo la metodología que se explicará en la sección siguiente CRISP-DM [Chapman *et al.*, 2000]). A pesar de ello, y por la gran importancia que tiene la información y la utilización de la

misma para proveer de conocimiento a las empresas para que puedan llevar a cabo la toma de decisiones oportunas y eficaces que brinden beneficios a la organización, es necesario tener en cuenta los riesgos que en este tipo de proyectos pueden llegar a ocurrir, para lograr estabilidad y que el proyecto no sufra contratiempos.

Es por ello que se detecta la necesidad de desarrollar un Modelo de Ciclo de Vida para este tipo de proyectos, el cual incluya un análisis y trato de los riesgos existentes y resulte eficaz para la empresa que vaya a llevar a cabo este tipo de Proyectos particulares pero usuales en estos tiempos. Además, resulta necesario la construcción de un Mapa de Actividades que determine qué actividades se llevan a cabo y cuales se descartan, y que tenga como ejes el Modelo de Ciclo de Vida a proponer y el Modelo de Procesos para Proyectos de Explotación de Información seleccionado [Vanrell, 2012]

1.4. Metodología empleada

Por medio de este trabajo se pretende lograr la construcción de un Modelo de Ciclo de Vida para Proyectos de Explotación de Información, que contemple los riesgos y, además, la construcción de un Mapa de Actividades que tenga como ejes en primer lugar al Modelo de Ciclo de Vida propuesto y en segundo lugar el Modelo de Procesos seleccionado y que se describirá en la sección 2.4.

En tal sentido, se describen los siguientes pasos metodológicos:

- La primera etapa de este trabajo consistió en la investigación de las distintas metodologías y herramientas para llevar a cabo Proyectos de Explotación de Información.
- Luego se analizaron los distintos Modelos de Ciclo de Vida para proyectos tradicionales de software y se describió cada uno, mostrando sus alcances y limitaciones.
- Posteriormente, y teniendo en cuenta las necesidades, se llevó a cabo la elección de un Modelo de Ciclo de Vida para Proyectos Tradicionales de Software en el cual se basará el Modelo de Ciclo de Vida a proponer.
- En cuarto lugar se describió la metodología más completa para desarrollar Proyectos de Explotación de Información y se identificaron sus alcances.
- Luego, se analizó y se describió el Modelo de Procesos seleccionado para realizar este trabajo y que tendrá un eje en el Mapa de Actividades a construir.
- En el siguiente paso se identificaron los procesos o fases de CRISP-DM [Chapman *et al.*, 2000] a incluir en el Modelo de Ciclo de Vida y se llevó a cabo una justificación y descripción de la elección

- A continuación se construyó el Modelo de Ciclo de Vida y posteriormente el Mapa de Actividades teniendo en cuenta ese Modelo de Ciclo de Vida y el Modelo de Procesos seleccionado en este trabajo.
- Se evaluó la viabilidad de la propuesta
- Finalmente se elaboró el informe final junto con las conclusiones obtenidas.

1.5. Estructura de este trabajo

Este trabajo se encuentra compuesto por 7 (siete) capítulos (cada uno compuesto por distintas secciones), los cuales se detallan a continuación:

En el capítulo 1 (uno) se presenta una Introducción al trabajo que se lleva a cabo, se establecen objetivos, tanto objetivos generales como específicos, se define el alcance que posee este trabajo, se evalúa los fundamentos que han dado soporte al desarrollo del mismo y posteriormente se describe la metodología empleada para lograr el desarrollo de este trabajo. El capítulo finaliza describiendo la estructura del mismo.

En el capítulo 2 (dos) se describe brevemente en que constan los Proyectos de Explotación de Información, al mismo tiempo se presentan los distintos tipos de Modelos de Ciclo de Vida que se encuentran para Proyectos de Desarrollo Tradicional de Software. Se selecciona un Modelo que sirve de base para el desarrollo del Modelo de Ciclo de Vida para Proyectos de Explotación de Información (el modelo de ciclo de vida en espiral de Boehm [1988].) y posteriormente, se analiza y describe en detalla cada fase de la metodología CRISP-DM que se utiliza para desarrollar Proyectos de este estilo. Al mismo tiempo se analiza COMPETISOFT el cual sirve de base para la construcción del Modelo de Procesos para Proyectos de Explotación de Información que se utiliza en este trabajo [Vanrell, 2012]. Finalizando este capítulo se tendrá una clara idea de las herramientas existentes y del Modelo de Procesos seleccionado

En el capítulo 3 (tres) nos centraremos en el problema existente de tener un Modelo de Ciclo de Vida para estos Proyectos de Explotación de Información, el cual sirva para el desarrollo de los mismos pero al mismo tiempo contemple y trate los riesgos. Se describe el problema en cuestión y al finalizar se realizan preguntas de investigación

En el capítulo 4 (cuatro) se propone una solución para el problema anteriormente descrito. Por lo tanto se lleva a cabo la construcción de un Modelo de Ciclo de Vida que contemple y tenga en cuenta los riesgos existentes en cualquier Proyecto de Explotación de Información. Al mismo tiempo se describe en detalle este Modelo propuesto. Seguido a esto se construye un Mapa de Actividades que tenga como ejes en primer lugar el Modelo de Ciclo de Vida propuesto y en

segundo lugar los procesos presentes en el Modelo de Procesos seleccionado en este trabajo [Vanrell, 2012]

Posteriormente, en el capítulo 5 (cinco) se presenta un caso de prueba que sirve de base para evaluar las propuestas realizadas y la viabilidad de este trabajo.

En el capítulo 6 (seis) se describen las conclusiones obtenidas a partir del desarrollo de este Trabajo, dando respuesta a los interrogantes de investigación y describiendo las futuras líneas de investigación a desarrollar a partir del mismo.

En el capítulo 7 (siete) se encuentra una lista de las referencias bibliográficas que se han utilizado en este Trabajo.

2. Estado del Arte

En esta sección se lleva a cabo una Introducción del Estado de la Cuestión (sección 2.1.). Posteriormente en la sección 2.2. se describen y analizan los Modelos de Ciclo de Vida que son de interés para el Trabajo Final (tanto el Modelo de Ciclo de Vida en Espiral de Boehm [1988]. como la metodología CRISP-DM [Chapman *et al.*, 2000]). Además, en la sección 2.3. se presentan los Modelos de Procesos que son de interés para este Trabajo Final: en primer lugar se lleva a cabo una descripción del Modelo de Procesos Competisoft (subsección 2.3.1.) describiendo las distintas categorías en las que se divide el modelo, Alta Dirección (subsección 2.3.1.1.), Gerencia (subsección 2.3.1.2) y Operación (subsección 2.3.1.3); y en segundo lugar se analiza y describe el Modelo de Procesos seleccionado por este trabajo (subsección 2.3.2.), el cual tiene como base el modelo de Procesos de Competisoft y las fases y actividades que se detallan en la metodología CRISP-DM para Proyectos de Explotación de Información.

2.1. Introducción

La Explotación de Información consiste en la extracción de conocimiento no-trivial que reside de manera implícita en los datos disponibles en distintas fuentes de información [Schiefer *et al.*, 2004]. Dicho conocimiento es previamente desconocido y puede resultar útil para algún proceso de decisión [Stefanovic *et al.*, 2006]. Para un experto, o para el responsable de un sistema de información, normalmente no son los datos en sí lo más relevante, sino el conocimiento que se encierra en sus relaciones, fluctuaciones y dependencias. Esta disciplina engloba un conjunto de técnicas de Minería de Datos (Data Mining, DM) encaminadas a la extracción de conocimiento procesable, implícito en el almacén de datos (Data Warehouse, DW) de la organización. Las bases de estas técnicas se encuentran en el análisis estadístico y en los sistemas inteligentes [Umopathy, 2007].

En los Proyectos de Explotación de Información se aborda, en general, la solución a problemas de clasificación, segmentación o identificación de atributos relevantes [Britos, 2008]. Estos resultados contribuyen con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003].

En [Sumathi y Sivanandam, 2006] se lleva a cabo un listado y descripción de las ventajas y beneficios que se obtienen a partir de la utilización de un proceso general de explotación de información:

- 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.

Algunas de las características comunes que tienen los proyectos de explotación de información fueron indicadas por Rabuñal Dopico en 2002. Estas características comunes 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.

La información es un patrimonio muy importante para todas las organizaciones y, la utilización y aprovechamiento de la misma de forma correcta y oportuna, le puede producir infinitos beneficios ya que por medios de procesos de explotación de información los miembros de la organización se puede dotar de conocimiento que resulta indispensable para la toma de decisiones, logrando certeza y eficiencia en las mismas.

A continuación se presentan tanto los Modelos de Ciclo de Vida que son de interés para este Trabajo Final, como así también mas adelante los Modelos de Procesos que resultan importantes para la realización de dicho trabajo.

2.2. Modelos de Ciclo de Vida de Interés para el Trabajo Final

En esta sección se describen los modelos de ciclo de vida que son de interés para la realización de este Trabajo Final. En primer lugar, en la subsección 2.2.1. se describen los distintos modelos de Ciclo de Vida existentes para Proyectos de Desarrollo Convencional de Software haciendo énfasis en el Modelo de Ciclo de Vida en Espiral de Boehm [1988], y en segundo lugar, en la subsección

2.2.2. se describe la metodología CRISP-DM presente para Proyectos de Explotación de Información

2.2.1. Modelo de ciclo de vida en espiral de Boehm

La Ingeniería de software se establece a partir de una serie de modelos que establecen y muestran las distintas etapas y estados por los que pasa un producto software, desde su concepción inicial, pasando por su desarrollo, puesta en marcha y posterior mantenimiento, hasta la retirada del producto. A estos modelos se les denomina modelos de ciclo de vida del software. El primer modelo concebido fue el de *Desarrollo en Cascada*. Este modelo se representa como un simple modelo con forma de cascada de las etapas del software. En este modelo la evaluación del producto software procede a través de una secuencia ordenada de transiciones de una fase a la siguiente según un orden lineal.

Otro modelo de ciclo de vida concebido es el de *Prototipado*, el cual se utiliza cuando el cliente no tiene una idea muy detallada de lo que necesita, o cuando el ingeniero de software o esta muy seguro de la viabilidad de la solución que tiene en mente. Este modelo tiene como objetivo contrarestar la congelación de requisitos mal comprendidos.

El desarrollo en *Espiral* [Boehm *et al.*, 1988] es un modelo de ciclo de vida del software definido por primera vez por Barry Boehm en 1986, utilizado generalmente en la Ingeniería de software. Las actividades de este modelo se conforman en una espiral, en la que cada bucle o iteración representa un conjunto de actividades. Las actividades no están fijadas a ninguna prioridad, sino que las siguientes se eligen en función del análisis de riesgo, comenzando por el bucle interior. Este modelo de ciclo de vida tiene en cuenta fuertemente el riesgo que aparece a la hora de desarrollar software. Para ello, se comienza mirando las posibles alternativas de desarrollo, se opta por la de riesgo más asumible y se hace un ciclo de la espiral. En cada vuelta o iteración hay que tener en cuenta:

- **Los Objetivos:** qué necesidad debe cubrir el producto.
- **Alternativas:** las diferentes formas de conseguir los objetivos de forma exitosa, desde diferentes puntos de vista como pueden ser: Características (experiencia del personal, requisitos a cumplir, etc.), Formas de gestión del sistema, Riesgo asumido con cada alternativa.
- **Desarrollar y Verificar:** Programar y probar el software.

Si el resultado no es el adecuado o se necesita implementar mejoras o funcionalidades:

- Se planificaran los siguientes pasos y se comienza un nuevo ciclo de la espiral. La espiral tiene una forma de caracol y se dice que mantiene dos dimensiones, la radial (Indica el aumento del coste del proyecto, ya que con cada nueva iteración se pasa más tiempo desarrollando) y la angular (Indica el avance del proyecto del software dentro de un ciclo):

Este sistema es muy utilizado en proyectos grandes y complejos como puede ser, por ejemplo, la creación de un Sistema Operativo. Al ser un modelo de Ciclo de Vida orientado a la gestión de riesgo se dice que uno de los aspectos fundamentales de su éxito radica en que el equipo que lo aplique tenga la necesaria experiencia y habilidad para detectar y catalogar correctamente los riesgos.

En la figura 2.1. se presenta la imagen del Modelo de Ciclo de Vida en Espiral, identificando fases, análisis, prototipos y las actividades de cada eje.

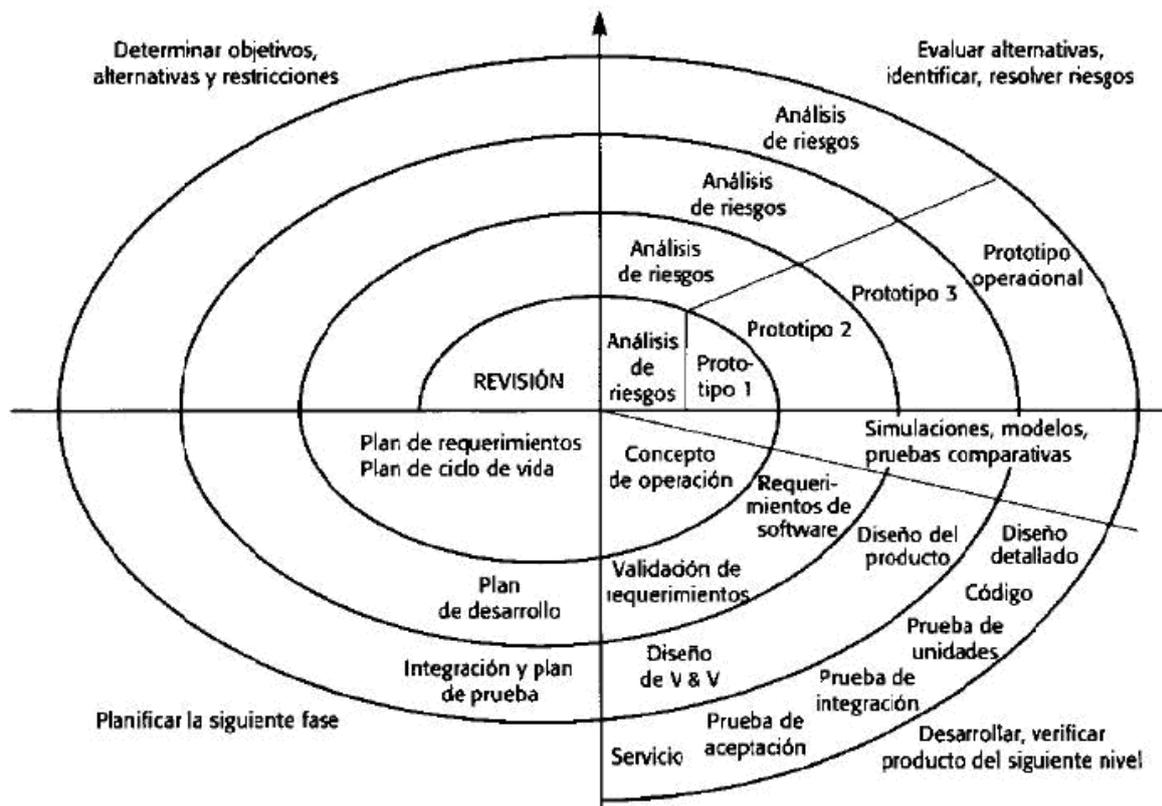


Figura 2.1. Modelo de Ciclo de Vida en Espiral de Boehm [1988]

Este modelo de Ciclo de Vida define cuatro actividades principales representadas por cuatro cuadrantes:

- Determinar Objetivos:** Fijar también los productos definidos a obtener: requerimientos, especificación, manual de usuario, Fijar las restricciones, Identificación de riesgos del proyecto y estrategias alternativas para evitarlos, Hay una cosa que solo se hace una vez: planificación inicial.
- Análisis del Riesgo:** Se lleva a cabo el estudio de las causas de las posibles amenazas y probables eventos no deseados y los daños y consecuencias que éstas puedan producir.
- Desarrollo y Testeo:** Desarrollo y verificación del producto del siguiente nivel
- Planificación:** Revisamos todo lo hecho, evaluándolo, y con ello decidimos si continuamos con las fases siguientes y planificamos la próxima actividad.

Las principales ventajas encontradas en este modelo de ciclo de vida en comparación con otros modelos son que:

- Su rango de opciones permiten utilizar los modelos de proceso de construcción de software tradicionales,
- Su orientación al riesgo evita muchas dificultades,
- Se presta atención a las opciones que permiten la reutilización de software existente,
- Se centra en la eliminación de errores y alternativas poco atractivas,
- No establece una diferenciación entre desarrollo de software y mantenimiento del sistema,
- Proporciona un marco estable para desarrollos integrados hardware-software,
- Incorpora objetivos de calidad e,
- Integra el desarrollo con el mantenimiento, etc.

Entre las desventajas se puede mencionar que:

- Genera mucho tiempo en el desarrollo del sistema,
- Es un modelo costoso y,
- Requiere experiencia en la identificación de riesgos.

2.2.2. CRISP-DM

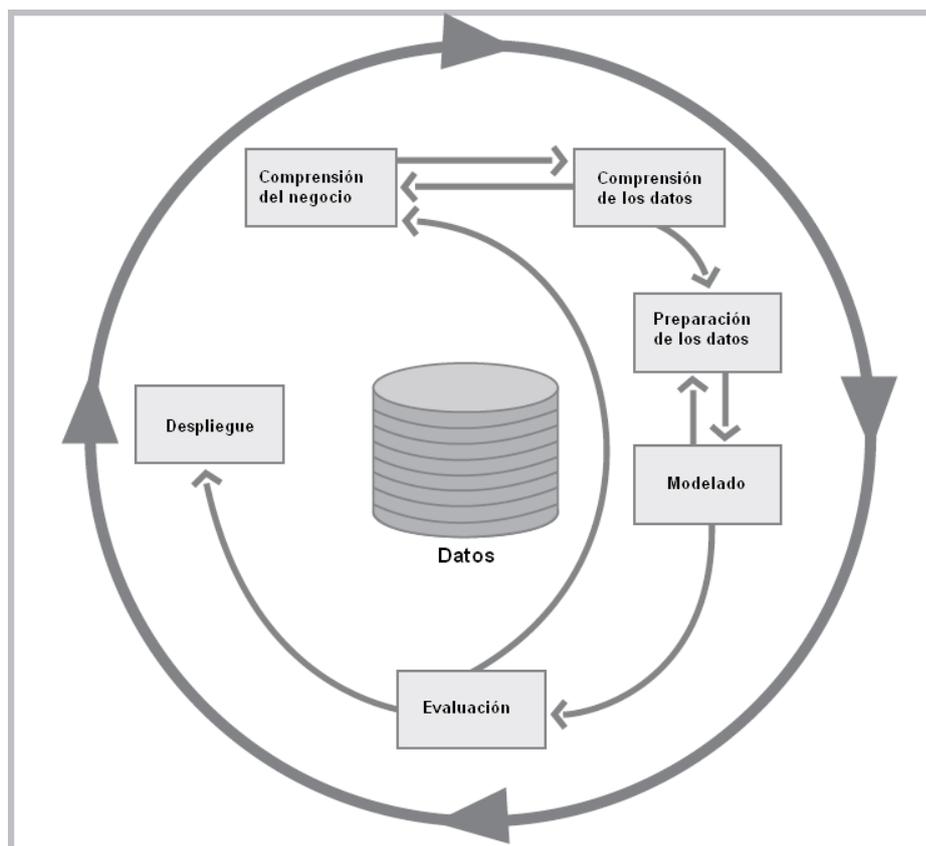
Para implementar una tecnología en un negocio, se requiere de una metodología. La mayoría de las consultoras especializadas en alguna tecnología cuentan, con por lo menos, una metodología, según los tipos de proyectos que aborden. Estos métodos son definidos a partir de sus experiencias y tomando lo mejor de los procedimientos más exitosos o populares. Contar con una metodología, se ha convertido tan importante y necesario como la carta de presentación de las empresas. Para los diferentes tipos de tecnologías, hay varias metodologías, algunas están publicadas en Internet. Para el caso de proyectos de implementación de minería de datos, hay una en particular: CRISP-DM [Chapman *et al.*, 2000].

Esta metodología define un ciclo de vida de los **proyectos de explotación de Información** que define las principales fases de un proyecto de este tipo. Estas fases son: Entendimiento de Negocios, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue.

En la imagen 2.2. se puede observar el diseño y las relaciones entre las distintas fases de esta metodología.

A continuación se analizará cada una de esas fases, indicando las actividades y propósitos generales de cada una de las mismas:

- Comprensión del negocio (Objetivos y requerimientos desde una perspectiva no técnica)
 - Establecimiento de los objetivos del negocio (Contexto inicial, objetivos, criterios de éxito)
 - Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio,...)
 - Establecimiento de los objetivos de la minería de datos (objetivos y criterios de éxito)
 - Generación del plan del proyecto (plan, herramientas, equipo y técnicas)
- Comprensión de los datos (Familiarizarse con los datos teniendo presente los objetivos del negocio)
 - Recopilación inicial de datos
 - Descripción de los datos
 - Exploración de los datos
 - Verificación de calidad de datos
- Preparación de los datos (Obtener la vista minable o dataset)
 - Selección de los datos
 - Limpieza de datos
 - Construcción de datos
 - Integración de datos
 - Formateo de datos
- Modelado (Aplicar las técnicas de minería de datos a los dataset)
 - Selección de la técnica de modelado
 - Diseño de la evaluación
 - Construcción del modelo
 - Evaluación del modelo
- Evaluación (De los modelos de la fase anteriores para determinar si son útiles a las necesidades del negocio)
 - Evaluación de resultados
 - Revisar el proceso
 - Establecimiento de los siguientes pasos o acciones
- Despliegue (Explotar utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización)
 - Planificación de despliegue
 - Planificación de la monitorización y del mantenimiento
 - Generación de informe final
 - Revisión del proyecto



• **Figura 2.2.** Metodología CRISP-DM [Chapman *et al.*, 2000]

El modelo de proceso jerárquico de CRISP-DM posee cuatro niveles distintos de abstracción: el nivel superior, el proceso de minería de datos es organizado en un número de fases; cada fase consiste de varias tareas genéricas de segundo nivel.

Este segundo nivel lo llaman genérico porque está destinado a ser bastante general para cubrir todas las situaciones posibles de minería de datos. Las tareas genéricas están destinadas a ser tan completas y estables como sea posible. Completo significa que cubre tanto al proceso entero de minería de datos y todas las aplicaciones de minería de datos posibles. Estable significa que el modelo debería ser válido para acontecimientos normales y aún para desarrollos imprevistos como técnicas de modelado nuevo.

El tercer nivel, el nivel de tarea especializado, es el lugar para describir como las acciones en las tareas genéricas deberían ser realizadas en ciertas situaciones específicas. Por ejemplo, en el segundo nivel podría haber una tarea genérica llamada limpieza de datos.

El tercer nivel describe como esta tarea se diferencia en situaciones diferentes, como la limpieza de valores numéricos contra la limpieza de valores categóricos, o si el tipo de problema es agrupamiento o el modelado predictivo. La descripción de fases y tareas como pasos discretos realizados en un orden específico representa una secuencia idealizada de eventos. En la práctica, muchas de las tareas pueden ser realizadas en una orden diferente, y esto a menudo será necesario

volver a hacer tareas anteriores repetidamente y repetir ciertas acciones. Nuestro modelo de proceso no intenta capturar todas estas posibles rutas del proceso de la minería de datos porque esto requeriría un modelo de proceso demasiado complejo.

El cuarto nivel, la instancia de proceso, es un registro de las acciones, decisiones, y de los resultados de una minería de datos real contratada.

En la imagen 2.3. se pueden ver los distintos niveles de abstracción de la metodología CRISP-DM para Proyectos de Explotación de Información y sus relaciones.

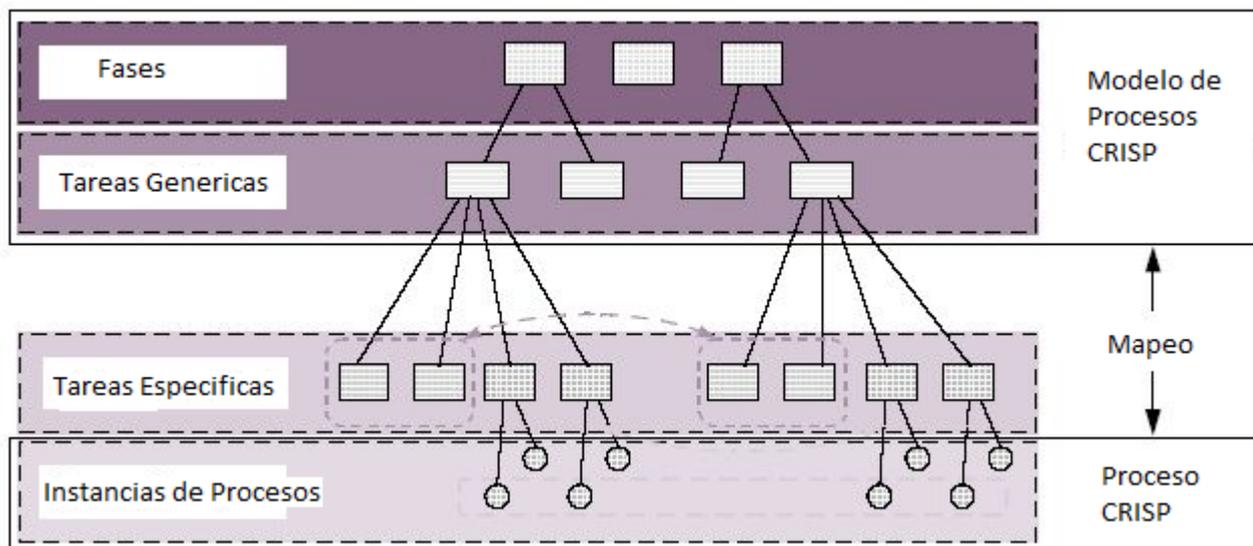


Figura 2.3. Niveles de abstracción de CRISP-DM [Chapman *et al.*, 2000]

Cada fase de esta metodología, cuenta con actividades específicas y documentación de salida asociada a cada una de estas tareas.

Teniendo como base la fase **Entendimiento del Negocio**, se puede observar en la tabla 2.1. sus actividades, descripción y documentación de salida correspondiente.

Por otro lado, en la tabla 2.2., se presentan los mismos rasgos para la fase **Entendimiento de los Datos** de CRISP-DM.

De igual manera, en la tabla 2.3., se lleva a cabo un listado de tareas y documentación de salida de la fase **Preparación de los Datos**.

En la tabla 2.4. se detallan las actividades y documentación de salida provenientes a la fase de **Modelado**.

Seguido a la fase de Modelado, se lleva a cabo la descripción de la fase **Evaluación** en la tabla 2.5.

Finalmente, en la tabla 2.6. se describe a la ultima fase de la metodología: **Despliegue**.

Esta metodología para proyectos de minería de datos es muy útil para comprender esta tecnología o extraer ideas para diseñar o revisar métodos de trabajo para proyectos de similares características.

Tarea	Descripción	Documentación de Salida
Determinación de los Objetivos del Negocio	Entendimiento, desde una perspectiva de negocio, de lo que el cliente realmente quiere lograr	Contexto (Background)
		Objetivos del Negocio
		Criterios del éxito del Negocio
Evaluación de la Situación	Investigación más detallada sobre todos los recursos, restricciones, presunciones, y otros factores que deberían ser considerados en la determinación del objetivo de análisis de datos y el plan de proyecto.	Inventario de Recursos
		Requerimientos, presunciones, y restricciones
		Riesgos y Contingencias
		Terminología
		Costos y Beneficios
Determinación de los objetivos de la minería de Datos	Un objetivo de negocio declara objetivos en la terminología de negocio. Un objetivo de minería de datos declara objetivos de proyecto en términos técnicos.	Objetivos de la minería de datos
		Criterios de éxito de la minería de datos
Producir el plan de proyecto	Describir del plan intencionado para alcanzar los objetivos de minería de datos y así alcanzar los objetivos de negocio	Plan del Proyecto
		Evaluación inicial de herramientas y técnicas

Tabla 2.1. Fase Entendimiento del Negocio

Tarea	Descripción	Documentación de Salida
Recoleccion Datos Iniciales	Adquiera en el proyecto los datos (o el acceso a los datos) listados en los recursos del proyecto.	Informe de colección de datos iniciales
Descripcion de los Datos	Examine las propiedades "gruesas" o "superficiales" de los datos e informe adquiridos en los resultados.	Informe de descripción de los datos
Exploracion de datos	Esta tarea dirige interrogantes de minería de datos usando preguntas, visualización, y técnicas de reporte	Informe de exploración de datos
Verificacion de calidad de datos	¿Los datos están completos? ¿Son correctos, o estos contienen errores y, si hay errores, que tan comunes son estos? ¿Hay valores omitidos en los datos?	Informe de calidad de datos

Tabla 2.2. Fase Entendimiento de los Datos

Tarea	Descripción	Documentación de Salida
Selección de datos	Decidir que datos serán usados para el análisis	Razonamiento para la exclusión/inclusión
Limpieza de datos	Elevar la calidad de los datos al nivel requerido por las técnicas de análisis.	Informe de la limpieza de datos
Construcción de datos	construcción de operaciones de preparación de datos tales como la producción de atributos derivados o el ingreso de nuevos registros	Atributos derivados
		Registros generados
Integración de datos	Métodos por el cual la información es combinada de múltiples tablas	Combinación de datos
Formateo de datos	Formateando transformaciones se refiere a modificaciones principalmente sintácticas hechas a los datos que no cambian su significado	Datos reformateados

Tabla 2.3. Fase Preparación de los Datos

Tarea	Descripción	Documentación de Salida
Escogimiento de la técnica de modelado	Selección de la técnica de modelado real que está por ser usado.	Técnicas de modelado
		Presunciones del modelado
Generación de la prueba de diseño	Generación de procedimiento o el mecanismo para probar la calidad y validez del modelo	Prueba de Diseño
Construcción del modelo	Ejecución la herramienta de modelado sobre el conjunto de datos preparados para crear uno o más modelos	Parámetros de ajustes
		Modelos
		Descripciones del modelo
Evaluación del modelo	Interpretación los modelos según su conocimiento de dominio, los criterios de éxitos de minería de datos, y el diseño de prueba deseado	Evaluación de modelos
		Parámetros de ajustes revisados

Tabla 2.4. Fase Modelado

Tarea	Descripción	Documentación de Salida
Evaluación de los resultados	Evalúa el grado al que el modelo responde (encuentra) los objetivos de negocio y procura determinar si hay alguna decisión de negocio por el que este modelo es deficiente	Evaluación de los resultados de la minería de datos en lo que concierne a criterios de éxito de negocio
		Modelos aprobados
Proceso de revisión	Los modelos resultantes pasan a ser satisfactorios y a satisfacer las necesidades de negocio.	Revisión del proceso
Determinación de los próximos pasos	El equipo de proyecto decide como proceder.	Lista de posibles acciones
		Decisión

Tabla 2.5. Fase Evaluación

Tarea	Descripción	Documentación de Salida
Desarrollo del plan	Esta tarea toma los resultados de la evaluación y determina una estrategia para el desarrollo	Desarrollo del plan
Plan de supervisión y mantenimiento	La supervisión y el mantenimiento son cuestiones importantes si los resultados de minería de datos son parte del negocio cotidiano y de su ambiente.	Supervisión y plan de mantenimiento
Producción de informe final	En el final del proyecto, el líder del proyecto y su equipo sobrescribe un informe final	Informe definitivo
		Presentación final
Revisión del proyecto	Evaluación de lo que fue correcto y lo que se equivocó, lo que fue bien hecho y lo que necesita para ser mejorado	Documentación de la experiencia

Tabla 2.6. Fase Despliegue

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. 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.

Finalmente, 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.

2.3. Modelos de Procesos de Interés para el Trabajo Final

En esta sección se describen los distintos Modelos de Procesos que son de interés para el Trabajo Final que se presenta. Esta misma sección se divide en 2 subsecciones: en primer lugar en la subsección 2.3.1. se analiza y describe al Modelo de Procesos CompetiSoft, y posteriormente en la subsección 2.3.2. se presenta el Modelo de Procesos propuesto por Vanrell [2012] en su Trabajo de Tesis.

2.3.1. Competisoft

En su trabajo, Vanrell [2012] lleva a cabo un profundo análisis del modelo de Procesos Competisoft, el cual utiliza y se basa para poder llevar a cabo la presentación de un Modelo de Procesos para Proyectos de Explotación de Información para la realización de su Tesis de Maestría. Él menciona que 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.

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”.

A continuación y tal como se lleva a cabo en el trabajo *Modelo de Proceso de Operación para Proyectos de Explotación de Información*, se analizará cada una de las categorías definidas en este Modelo de Procesos: en la sección 2.3.1. se describirá a la Alta Dirección, en la sección 2.3.2. se analizará a la Gerencia y en la sección 2.3.3. se describirán las funciones del nivel Operativo.

2.3.1.1. 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.

2.3.1.2. 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.

2.3.1.3. 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.

2.3.2. Modelo de Procesos

En su trabajo *Modelo de Proceso de Operación para Proyectos de Explotación de Información* [Vanrell, 2012], Vanrell propone un marco teórico para la creación de modelo de procesos de operación para proyectos de explotación de información para PYMEs, siguiendo los lineamientos del modelo de procesos para la industria de software (COMPETISOFT). Este trabajo tiene como objetivos específicos:

- 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.

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, Vanrell 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.

Como problema, en el trabajo mencionado se menciona que 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.

Al mismo tiempo, el autor al llevar a cabo una evaluación de las principales metodologías existentes para los proyectos de explotación de información mencionados anteriormente, observó la falta de herramientas que permitían soportar de forma completa la fase de administración de proyectos.

Por lo tanto, y teniendo en cuenta los problemas surgidos, el trabajo mencionado en esta sección se orientó 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 su trabajo propone mantener los procesos definidos en COMPETISOFT como Administración de Proyectos y Desarrollo de Proyectos readecuándolos a los proyectos de explotación de información. Para esto elimina las tareas del proceso de desarrollo y las reemplaza por las fases de CRISP-DM que se relacionan con el desarrollo del proyecto.

La reestructuración de estos dos procesos en el desarrollo de proyectos de Explotación de Información brinda 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. Así surgen 2 procesos: uno vinculado a la administración de proyectos de explotación de información que se muestra en la tabla 2.7 (a y b) y otro relacionado con el desarrollo de proyectos de explotación de información que se puede observar en la tabla 2.8. En estas tablas se pueden observar para cada subproceso tanto de Administración de Proyecto como Desarrollo del Procesos las tareas a llevar a cabo y desarrollar, como así también la salida correspondiente

Las actividades de administración de proyectos son actividades que deben llevarse en paralelo en procesos separados. Es por esto que el autor decidió separar algunas de las tareas definidas en el ciclo de desarrollo de CRISP para incluirlas en un ciclo diferente que pasará a formar parte del proceso de administración de proyectos. Esto sucede con el ciclo definido en CRISP como Entendimiento de Negocio que incorpora las tareas para obtener un ida y vuelta con la organización para la cual se va a realizar el proyecto, la de definir los objetivos del negocio y la definición de los criterios de éxito del negocio, también la definición de un plan de entrega que en COMPETISOFT se tiene en cuenta dentro de la fase de definición del protocolo de entrega con el cliente, la definición de requerimientos del proyecto, suposiciones y restricciones, la evaluación de riesgos (riesgos detectado se mencionaran en la sección siguiente) y construcción de planes de contingencia, la construcción de una base con terminología de negocio y la definición de costos y

beneficios que traerá el proyecto a la organización al igual que la producción de un plan de proyecto y un plan de desarrollo. Las fases definidas para los proyectos de desarrollo clásico no son representativas de las fases naturales de un proyecto de explotación de información las cuales están completamente relacionadas con las distintas transformaciones que sufren los datos a lo largo del desarrollo del proyecto. Estas fases son las denominadas por CRISP como:

- Entendimiento de Negocio,
- Entendimiento de los Datos,
- Preparación de los Datos,
- Modelado,
- Evaluación. y
- Entrega/Despliegue.

A continuación se presenta el modelo de procesos propuesto por Vanrell [2012], teniendo en cuenta las tareas a desarrollar por cada subproceso mencionado, como así también la/s salida/s correspondiente a cada tarea, mencionando posteriormente las distintas técnicas que se podrían aplicar a las tareas definidas. Como ya se mencionó, este modelo de procesos está dividido en subprocesos para Administración de Proyectos y en subprocesos para llevar a cabo el Desarrollo del Proyecto. En las tablas 2.7.a) y 2.7.b) se presenta los subprocesos de Administración de Proyectos, los cuales tienen como función recolectar información que le sea útil a los procesos de Desarrollo para aumentar la calidad de los mismos permitiendo realizar ajustes y mantener un estándar en la realización de distintos proyectos llevando a cabo la conversión de procesos caóticos y aislados en procesos repetitivos y controlados.

Seguido a la descripción de los subprocesos de Administración del Proyecto, en la tabla 2.8 se definen las tareas y salidas correspondientes a los subprocesos de Desarrollo del Proyecto, los cuales poseen como fin “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”. Para cada tarea, en el trabajo se adhirieron una serie de técnicas que pueden ser aplicadas como recomendación para llevar a cabo las actividades que se mencionan en el proceso de administración de proyectos.

A continuación se detallan algunas de ellas:

- 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 ha mencionado en secciones anteriores, naturalmente las fases de desarrollo de los proyectos de Explotación de Información no coinciden con las fases mediante las cuales se desarrollan los proyectos de software clásicos.

Por ello, los subprocesos de desarrollos fueron adaptados teniendo en cuentas las fases de la metodología CRISP-DM para llevar a cabo Proyectos de Explotación de Información.

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 conclusión de este trabajo descripto, Vanrell [2012] considera que se ha logrado construir un Modelo de Procesos para Proyectos de Explotación de Información que contempla bien diferenciados los procesos de Administración y los procesos de Desarrollo.

Los procesos de Administración que fue el valor agregado que le dio Vanrell [2012] a la metodología existente para este tipo de Proyectos (CRISP-DM) tienen como fin “mantener los

Procesos de Desarrollo en movimiento y corregir aquellos desvíos que se produzcan con el fin de lograr una finalización exitosa del proyecto”.

Subproceso	Tarea	Salida
Planificación /Entendimiento del Negocio	Entendimiento del Negocio	<ul style="list-style-type: none"> • Conocimiento del negocio Objetivos del negocio • Criterios de éxito
	Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento	<ul style="list-style-type: none"> • Proceso Especifico (forma parte del Plan de Desarrollo)
	Definir el protocolo de entrega con el cliente	<ul style="list-style-type: none"> • Plan de Entrega
	Definir ciclos y actividades con base en la descripción del proyecto y en el proceso específico	<ul style="list-style-type: none"> • Proceso Especifico (forma parte del Plan de Desarrollo)
	Determinar tiempo estimado para cada actividad	<ul style="list-style-type: none"> • Calendario de actividades (forma parte del Plan de Desarrollo) incorpora el tiempo estimado en el Plan de Proyecto
	Elaborar plan de adquisiciones y capacitación	<ul style="list-style-type: none"> • Plan de Adquisiciones y Capacitación
	Establecer el equipo de trabajo	<ul style="list-style-type: none"> • Equipo de trabajo (forma parte del Plan de Desarrollo)
	Establecer el calendario de actividades	<ul style="list-style-type: none"> • Calendario de actividades (forma parte del Plan de Desarrollo)
	Calcular el costo estimado del proyecto	<ul style="list-style-type: none"> • Costo estimado
	Evaluación de la situación	<ul style="list-style-type: none"> • Inventario de recursos Requerimientos, suposiciones y restricciones • Riesgos y • Terminología • Costos y beneficios
	Producir un Plan de Proyecto	<ul style="list-style-type: none"> • Plan de Proyecto, incluye ciclos y actividades, tiempo estimado, plan de adquisiciones y capacitación, equipo de trabajo, costo estimado, calendario, plan de manejo de riesgos y protocolo de entrega
	Producir un Plan de Desarrollo	<ul style="list-style-type: none"> • Plan de Desarrollo (incluye descripción del producto y entregables, proceso específico, equipo de trabajo y calendario) Lista inicial de técnicas y herramientas
	Formalizar el inicio de un nuevo ciclo del proyecto	

Tabla 2.7. a) Procesos de Administración

ubproceso	Tarea	Salida
Realización	Acordar las tareas con el equipo de trabajo	
	Acordar la distribución de información	
	Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario	
	Revisar cumplimiento del plan de adquisiciones y capacitación	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Administrar subcontratos	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento Reporte de Mediciones y Sugerencias de Mejora
	Registrar costo real del proyecto	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Revisar el registro de rastreo basado en los productos de trabajo recolectados	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Revisar los productos terminados durante el Proyecto	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Recibir y analizar las solicitudes de cambio del cliente	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Evaluación y Control	Evaluar el cumplimiento del plan de proyecto y plan de desarrollo	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Analizar y controlar los riesgos	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
	Generar el reporte de seguimiento del proyecto	<ul style="list-style-type: none"> Reporte de Seguimiento / Plan de monitoreo y mantenimiento
Cierre/ Entrega	Formalizar la terminación del proyecto o ciclo	<ul style="list-style-type: none"> Documento de aceptación
	Llevar a cabo el cierre del contrato con Subcontratistas	
	Generar el reporte de mediciones y sugerencias de mejora	<ul style="list-style-type: none"> Reporte de mediciones y sugerencia de mejoras – Lecciones Aprendidas
	Planear la entrega	<ul style="list-style-type: none"> Plan de entrega (forma parte del Plan de Proyecto nombrado como protocolo de entrega)

Tabla 2.7. b) Procesos de Administración

Este proceso de Administración tiene la función de recolectar información que le sea útil a los procesos de Desarrollo para aumentar la calidad de los mismos permitiendo realizar ajustes y mantener un estándar en la realización de distintos proyectos llevando a cabo la conversión de procesos caóticos y aislados en procesos repetitivos y controlados.

Los procesos de Desarrollo poseen como fin “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”.

Subproceso	Tarea	Salida
Entendimiento del Negocio	Determinar las metas del Data Mining	<ul style="list-style-type: none"> Metas del Data Mining Criterios del éxito del Data Mining
Entendimiento de los Datos	Reunir los datos iniciales	<ul style="list-style-type: none"> Reporte de datos iniciales
	Describir los datos	<ul style="list-style-type: none"> Reporte de descripción de los datos
	Explorar los datos	<ul style="list-style-type: none"> Reporte de exploración de datos
	Verificar la calidad de los datos	<ul style="list-style-type: none"> Reporte de calidad de los datos
Preparación de los datos	Tareas preparatorias	<ul style="list-style-type: none"> Datasets Descripción de los datasets
	Seleccionar los datos	<ul style="list-style-type: none"> Justificación de inclusión / exclusión
	Limpiar los datos	<ul style="list-style-type: none"> Reporte de limpieza de datos
	Construir los datos	<ul style="list-style-type: none"> Atributos derivados
		<ul style="list-style-type: none"> Registros generales
	Integrar los datos	<ul style="list-style-type: none"> Datos combinados (combinación de tablas y agregaciones)
Modelado	Formatear los datos	<ul style="list-style-type: none"> Datos formateados
	Seleccionar las técnicas de modelado	<ul style="list-style-type: none"> Técnica de modelado Suposiciones de modelado
	Generar el diseño de test	<ul style="list-style-type: none"> Diseño del test
	Construir el modelo	<ul style="list-style-type: none"> Establecimiento de parámetros Modelos Descripción del modelo
	Evaluar el modelo	<ul style="list-style-type: none"> Evaluación del modelo
Evaluación		<ul style="list-style-type: none"> Revisión de los parámetros establecidos
	Evaluar resultados	<ul style="list-style-type: none"> Evaluación de los resultados de Data Mining respecto a los criterios del éxito Modelo aprobado
	Revisar el proceso	<ul style="list-style-type: none"> Revisión del proceso
Entrega	Determinar próximos pasos	<ul style="list-style-type: none"> Lista de posibles decisiones Decisiones
	Producir un reporte final	<ul style="list-style-type: none"> Reporte final Presentación final

Tabla 2.8. Procesos de Desarrollo

Teniendo en cuenta todo lo presentado y aprendido en este capítulo, en el siguiente se presenta y describe el problema en cuestión, para que posteriormente en el capítulo 4 se lleve a cabo la presentación de una solución propuesta a dichos problemas mencionados acompañada de una descripción de la misma.

3. Problema

En este capítulo se desarrolla la problemática que intenta solucionar este Trabajo (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 Final de Licenciatura (sección 3.3)

3.1. Contexto del Problema

Al intentar llevar adelante diversos Proyectos de Explotación de Información se utilizan diferentes metodologías y herramientas para poder llevar a cabo el desarrollo de los mismos. Estos proyectos a diferencia de los proyectos tradicionales poseen características únicas, por lo cual las herramientas disponibles no se adecuan completamente a los Proyectos de E.I.

En la actualidad más allá de proyectos de software tradicional cada día más empresas desarrollan proyectos de explotación de información que les permitirá desarrollar conocimiento el cual beneficiará al momento de la toma de decisiones, por lo que es necesario metodologías y herramientas que puedan estimarlos, analizando riesgos existentes en los mismos y mejorando su calidad. Este trabajo apunta a eso.

3.2. Descripción del problema

Teniendo en cuenta y, aceptando que los proyectos de Explotación de Información poseen características particulares y distintas a los proyectos convencionales y tradicionales de software sobre todo en la parte operativa de un proyecto [Vanrell *et al.*, 2010a], es necesario contar con un Modelo de Ciclo de Vida para este tipo de Proyectos que se adapte a las necesidades y requerimientos de los mismos.

Teniendo como base las metodologías y modelos de ciclos de vida para proyectos tradicionales de software y contando con la metodología CRISP-DM para Proyectos de Explotación de Información, se ha encontrado que no hay ninguna herramienta o metodología que tenga en cuenta los posibles riesgos que pueden tener este tipo de proyectos. Los posibles riesgos identificados para proyectos de Explotación de Información cuentan con conceptos a ser educados [Britos, 2008]

Los riesgos identificados con cada concepto a ser educado se pueden ver en la tabla 3.1 presentada a continuación:

Problema	Conceptos a ser educados
a) El cliente no entiende el léxico técnico utilizado. b) Se puede no entender el léxico utilizado por el cliente c) Se puede encontrar difícil entender cómo ayudar al cliente ya que no se conoce el dominio del proyecto	Definiciones, acrónimos y abreviaturas
d) El cliente puede no estar seguro de lo que el proyecto puede hacer y lograr e) Los modelos definidos pueden ser distintos a los que el cliente esperó	Objetivos, criterios de éxito, expectativas y suposiciones del proyecto
f) El cliente está en un grupo impredecible (no tiene tanto interés por el proyecto)	Recursos humanos involucrados
g) El cliente no sabía las necesidades de información de la organización	Restricciones del proyecto, riesgos del proyecto y planes de contingencia
h) Los datos identificados por los requisitos no eran los correctos	Los requisitos de información o fuente de datos y los atributos relacionados con los requisitos
i) Cuando se está en fase de modelado y se detectan problemas en los datos, es necesario redefinir los requerimientos	Requerimientos de supuestos resultados, de restricciones, de riesgos y del plan de contingencia
j) Requisitos malentendidos del proyecto resultaron en la selección de la herramienta de modelado equivocada	Evaluación de herramientas de DM-BI

Tabla 3.1. Relación entre problemas/riesgos identificados y conceptos a ser educados

Para poder llevar a cabo una solución de los problemas presentados en la primera columna de la tabla anterior, es necesario analizar y describir los conceptos mencionados en la segunda columna:

- *Definiciones, acrónimos y abreviaturas:* Es necesario identificar definiciones, acrónimos y abreviaturas para establecer el léxico que se repartirán entre todas las personas relacionadas con el Proyecto.
- *Objetivos del Proyecto:* Es necesario identificar el objetivo del proyecto DM-BI y su motivación para caracterizar lo que el cliente necesita.
- *Criterios del éxito:* Es necesario identificar los criterios que convierten AL proyecto en un éxito. Los criterios deben ser descriptos en términos de logros esperados del Proyecto.

- *Expectativas del Proyecto:* Es necesario identificar lo que se espera que sea alcanzado por el Proyecto y para confirmar que cumplen con las expectativas del cliente. Las expectativas deben estar alineados con los objetivos y los criterios de éxito del proyecto.
- *Suposiciones del Proyecto:* Es importante identificar las suposiciones que deben ser asumidas como verdaderas. Las suposiciones del Proyecto se convierten en el punto de partida para el mismo.
- *Restricciones del Proyecto:* Con el fin de especificar el contexto del Proyecto, es necesario identificar los límites previamente establecidos para el relacionado con la organización.
- *Atributos relacionados con los requisitos:* Establecer los atributos que se va a utilizar con el fin de lograr un requisito objetivo.
- *Riesgos en requisitos:* Es importante identificar los riesgos de los requisitos del proyecto mirando continuamente por que podría estar mal el requisito (en relación con el proyecto) y determinar qué riesgos son importantes por resolver. La identificación de riesgos en los requisitos es necesaria para definir los planes de contingencia que se aplicarán cuando sea necesario.
- *Plan de Contingencia:* Es necesario definir los planes de contingencia que se aplicarán cuando una complicación ocurre.
- *Evaluación de herramientas de DM-BI:* Es necesario evaluar las herramientas disponibles para establecer que son las mejores para lograr los objetivos del proyecto.

La diferencia más significativa que se identifica entre los Proyectos de Explotación de Información y los Proyectos de Software Tradicional se presenta en los procesos de desarrollo y mantenimiento de software en los cuales Competisoft (modelo de procesos para proyectos tradicionales de software) 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 [Vanrell, 2012].

Teniendo en cuenta este problema de no contar con un Modelo de Ciclo de Vida que sirva para Proyectos de Explotación de Información (teniendo como base las fases descritas en CRISP-DM necesarias para llevar a cabo este tipo de Proyectos), que tenga en cuenta y analice los riesgos que pueden en ellos ocurrir; se detecta que es necesario construir un Modelo de Ciclo de Vida que tenga en cuenta los riesgos y sirva como base para el diseño de estos particulares estilos de proyectos y al mismo tiempo, diseñar a partir del Modelo desarrollado y del Modelo de Proceso seleccionado y descrito, un Mapa de Actividades donde se marcan qué actividades del proceso software se va a ejecutar en un determinado tiempo.

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:

- Pregunta 1: ¿Es posible construir un Modelo de Ciclo de Vida para Proyectos de Explotación de Información que además de las tareas necesarias para los mismos tenga en cuenta los riesgos?
- Pregunta 2: ¿Qué riesgos pueden tener este tipo de Proyectos a diferencia de Proyectos de Desarrollo de Software Tradicional?
- Pregunta 3: ¿Se podrá construir un Mapa de actividades teniendo como ejes el Modelo de Ciclo de Vida a proponer con el Modelo de Procesos descrito en la sección anterior que tenga en cuenta tanto los procesos de Administración como los de Desarrollo?

Luego de presentar estos interrogantes y teniendo en cuenta la problemática descrita en este capítulo, en el siguiente capítulo se presenta la propuesta de un Modelo de Ciclo de Vida adaptado a los Proyectos de Explotación de Información que solucione esta problemática existente y, posteriormente, se lleva a cabo la construcción de un Mapa de Actividades teniendo como ejes el Modelo de Procesos propuesto por Vanrell [2012] y el Modelo de Ciclo de Vida desarrollado en el presente Trabajo.

4. Solución

En este capítulo se desarrolla la solución propuesta y planteada para la problemática descrita anteriormente en el capítulo 3. En primer lugar se describe la propuesta (sección 4.1.) y luego se lleva a cabo una descripción de la solución basada en la propuesta (sección 4.2.). Dentro de la solución basada en la propuesta se presenta el Modelo de Ciclo de Vida presentado como solución (sección 4.2.1.) con subsecciones que describen cada vuelta que realiza el espiral y un breve análisis de la división por fase del Modelo de Procesos seleccionado (subsecciones 4.2.1.1. a 4.2.1.7.). Finalmente se construye el Mapa de Actividades teniendo como ejes el Modelo de Procesos seleccionado para este tipo de Proyectos y las fases del Modelo de Ciclo de Vida presentado (sección 4.2.2.) y luego se lleva a cabo un análisis detallado de las actividades desarrolladas por cada fase del Modelo de Ciclo de Vida (4.2.2.1. a 4.2.2.7.).

4.1. Propuesta

Dados los problemas mencionados y descritos en la sección anterior, se propone la creación de un Modelo de Ciclo de Vida en Espiral, con la esencia del Modelo creado por Boehm, que contemple y tenga un fuerte énfasis en los riesgos que en los Proyectos puedan ocurrir, y que sirva, y pueda ser utilizado, para los distintos Proyectos de Explotación de Información que realicen las empresas y organizaciones.

Para la realización de este Modelo de Ciclo de Vida se han utilizado y tomado las distintas actividades que se han descrito anteriormente en el Modelo de Procesos seleccionado para este tipo de Proyectos [Vanrell,2012]. Como se ha mencionado en el Estado del Arte, esta Metodología no contempla ni tiene en cuenta los riesgos que en un Proyecto de Explotación de Información pueden ocurrir. Los riesgos, algunos de los cuales fueron descritos en la sección 3 de este trabajo, pueden presentarse y sin un trato ni un análisis adecuado pueden provocar grandes costos dentro de las organizaciones.

Estos análisis son necesarios para que el desarrollo del Proyecto vaya por causas normales. Por eso mismo, el Modelo que se presentará cuenta en cada ciclo del Espiral con un Análisis de Riesgos, facilitando y asegurando que la siguiente fase no contará con inconvenientes previos.

4.2. Solución basada en la propuesta

Teniendo en cuenta el Problema descrito en el capítulo anterior, se lleva a cabo la presentación en primer lugar de un Modelo de Ciclo de Vida (sección 4.2.1.) y posteriormente la construcción de un

Mapa de Actividades (sección 4.2.2.) teniendo como ejes el Modelo de Ciclo de Vida planteado y el Modelo de Procesos seleccionado en este trabajo.

4.2.1. Modelo de Ciclo de Vida Propuesto

Como ya se ha mencionado, se ha tomado la base del Modelo de Ciclo en Espiral de Boehm readaptando las fases a las actividades descritas por el Modelo de Procesos desarrollado por Vanrell [2012] para Proyectos de Explotación de Información, el cual posee un valor agregado en comparación a la metodología existente para este tipo de Proyectos como es CRIS-DM. Estas actividades corresponden a las 6 fases de dicha metodología: Entendimiento del Negocio, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue.

Básicamente, en cada vuelta del espiral se realizan actividades correspondientes al Modelo de Procesos correspondiente para Proyectos de E.I. [Vanrell, 2012], luego se planea la próxima iteración, se lleva a cabo un análisis de riesgo y por último se lleva a cabo la presentación y desarrollo de un prototipo a partir de las fases de Desarrollo en si del Modelo de Ciclo de Vida. Estas actividades se realizan hasta cubrir todas las fases de la metodología para Proyectos de Explotación de Información.

Teniendo en cuenta la arquitectura y el diseño del Modelo de Ciclo en Espiral de Boehm, el Modelo propuesto mantiene los 4 ejes del espiral, los cuales son:

- Determinación de objetivos, alternativas, restricciones;
- Evaluación de alternativas. Identificación, resolución de riesgos;
- Desarrollo, verificación del producto del próximo nivel;
- Planificación de las fases.

Entre los ejes Evaluación de alternativas, identificación y resolución de riesgos (análisis de riesgos) y el eje de Desarrollo, verificación del producto del próximo nivel se desarrollan distintos prototipos y simulaciones, modelos y benchmark, tanto para ir viendo con el cliente si lo entendido por el grupo de desarrollo es lo correcto y se está avanzando adecuadamente en el transcurso del proyecto o si hay que modificar algo, como también para ir analizando el rendimiento del sistema y/o de los componentes desarrollados hasta el momento. Para este aspecto resulta necesaria la participación del cliente o de algún usuario final que vaya a utilizar lo desarrollado en el Proyecto y que haya definido los requisitos del mismo.

Teniendo en cuenta lo dicho, en la figura 4.1. se puede observar el Modelo de Ciclo de Vida desarrollado para cubrir las necesidades planteadas en la sección 3. Este modelo planteado incluye más vueltas al espiral para abarcar todos los procesos del Modelo de Procesos que derivan de las

fases de CRISP-DM y para garantizar la no aparición de riesgos y el desarrollo tranquilo del Proyecto.

Se ha modelado, en el Modelo de Ciclo de Vida presentado, a los procesos y tareas detectadas en el Modelo de Procesos seleccionado para Proyectos de Explotación de Información. Estas tareas y actividades son mencionadas por Vanrell como actividades o fases de los procesos de Desarrollo del Proyecto. Adicionalmente, y agregándole el valor que le ha dado y la mejora que ha hecho Vanrell a la metodología ya existente para este tipo de Proyectos, se ha determinado que es necesario ilustrar en el Modelo de Ciclo de Vida en Espiral propuesto las tareas y subprocesos correspondientes a la Administración del Proyecto [Vanrell,2012]. Estos subprocesos abarcan todo el largo del desarrollo del Proyecto y dan soporte a las actividades de desarrollo propiamente dichas.

Como se ha mencionado en el capítulo 2, tienen como fin “mantener los Procesos de Desarrollo en movimiento y corregir aquellos desvíos que se produzcan con el fin de lograr una finalización exitosa del proyecto” [Vanrell, 2012]. Este proceso de Administración que se ubica a lo largo del Modelo de Ciclo de Vida propuesto tiene la función de recolectar información que le sea útil a los procesos de Desarrollo para aumentar la calidad de los mismos permitiendo realizar ajustes y mantener un estándar en la realización de distintos proyectos llevando a cabo la conversión de procesos caóticos y aislados en procesos repetitivos y controlados.

Una vez explicado el porqué del diseño de esa manera de los SubProcesos correspondiente a las tareas de Administración del Proyecto, se pasa a explicar cada iteración y vuelta del Modelo de Ciclo de Vida propuesto.

4.2.1.1. Primera Vuelta

En la primera vuelta del espiral propuesto, en el eje de desarrollo (verificación del producto del próximo nivel) se lleva a cabo la tarea de determinar los Objetivos del Negocio. Mediante esta actividad, se describen los objetivos primarios del cliente, desde una perspectiva del negocio. Además de los objetivos del negocio primario, allí hay típicamente otras preguntas de negocio relacionadas con lo que al cliente le gustaría administrar. Por otro lado, se describen los criterios del éxito del negocio, los resultados útiles que se desean para el proyecto y bajo que circunstancias se va a dar como aprobado el desarrollo del proyecto a realizar. Pasando al próximo eje del espiral, se realiza la Planificación de Requisitos y Planificación del Ciclo de Vida del Proyecto de Explotación de Información a realizar.

Teniendo en cuenta la primera iteración, se lleva a cabo un análisis de riesgo (AR1), donde se lleva a cabo el estudio y previsión de futuros riesgos que puedan llegar a ocurrir teniendo en cuenta lo descripto, modelado y desarrollado hasta el momento.

En el Capítulo 3 se han mencionado una serie de Riesgos que un Proyecto de Explotación de Información puede llegar a tener y sus conceptos a ser educados [Britos, 2008].

Unos de esos Riesgos que se pueden llevar a cabo y se pueden analizar en esta parte del Espiral son tanto el del que el cliente puede no estar seguro de lo que el Proyecto puede hacer y lograr como así también que los modelos definidos pueden ser distintos a los que el cliente esperaba. Ambos riesgos o problemas identificados implican la necesidad de revisar los Objetivos del Negocio, Criterios del Éxito, Expectativas y Suposiciones del Proyecto.

Por otro lado se deben analizar los requisitos ya que puede ocurrir el riesgo de que ocurran y existan requisitos mal entendidos del proyecto los cuales a futuro pueden ocasionar la mala selección de la herramienta de modelado a utilizar para llevar a cabo el mismo. En la tabla 4.1. se representa lo anteriormente dicho: la relación entre el proceso del Espiral y sus correspondientes posibles riesgos a detectar en el análisis de Riesgo 1 (AR1).

Proceso del Espiral	Riesgos – Análisis Riesgo 1
Objetivos del Negocio	El cliente puede no estar seguro de lo que el proyecto puede hacer y lograr.

Tabla 4.1. Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR1

4.2.1.2. Segunda Vuelta

En la segunda vuelta del espiral, situándonos en el eje de Desarrollo, Adaptación y Pruebas, se continúan con las actividades de la fase de **Entendimiento del Negocio** de la metodología CRISP-DM y tomadas en los Procesos de Desarrollo de Proyectos del Modelo de Procesos de Vanrell [2012] para Proyectos de Explotación de Información. En primer lugar se lleva a cabo la Evaluación de la Situación, en la cual se realiza una investigación más detallada sobre los recursos, restricciones, presunciones, y otros factores que deberían ser considerados en la determinación del objetivo de análisis de datos y el plan de proyecto.

Las salidas de esta actividad serán Inventarios de recursos, Requerimientos, presunciones y restricciones, Riesgos y contingencias, Terminología y Costos y beneficios.

Posteriormente a esta actividad y siguiendo analizando la segunda vuelta del espiral propuesto en la figura 4.1. se tiene la Determinación de los Objetivos del Proyecto de Explotación de Información en sí. A diferencia de los Objetivos del Negocio planteados en la primera vuelta, en esta actividad se declaran los objetivos del proyecto en términos técnicos.

Como salida a esta actividad se obtienen los Objetivos del Proyecto de Explotación de Información (descripción de las salidas intencionadas del proyecto que permiten el logro de los objetivos de

negocio) y los Criterios de Éxito del Proyecto de Explotación de Información (definición de los criterios de un resultado exitoso para el proyecto en términos técnicos).

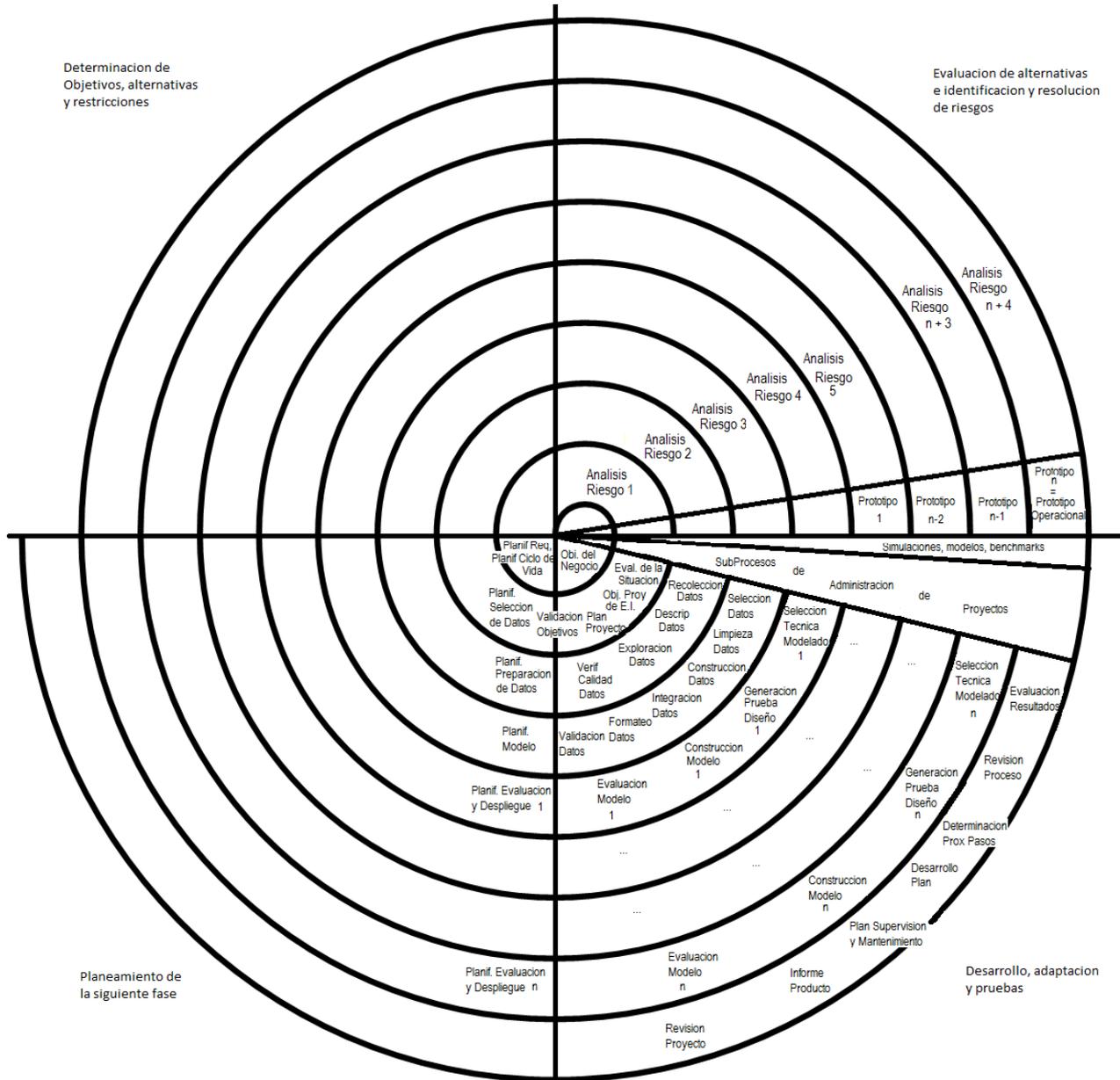


Figura 4.1. Modelo de Ciclo de Vida Propuesto

Luego de la descripción de los Objetivos del Proyecto de Explotación de Información se debe desarrollar un Plan de Proyecto, en el cual se especifican los pasos a ser realizados durante el resto del proyecto, incluyendo la selección inicial de herramientas y técnicas. Las salidas asociadas a esta actividad son el Plan de Proyecto en sí, donde se listan las etapas a ser ejecutadas en el proyecto, juntos con su duración, recursos requeridos, entradas, salidas, y dependencias; y además se tiene

como salida la Evaluación de Herramientas y Técnicas ya que es importante evaluar herramientas y técnicas temprano en el proceso ya que esto puede influir en el proyecto entero.

Finalizando con las tareas de la segunda vuelta del espiral y siguiendo situados en el eje de Desarrollo, Adaptación y Pruebas, se realiza al final una Validación de los Objetivos, donde se analizan los objetivos planteados para el Proyecto de Explotación de Información y se determina si cumple técnica y funcionalmente con lo que el cliente ha solicitado. Superando esta actividad, se pasa al eje de Planeamiento de la Siguiete Fase, donde se realiza la Planificación de la Selección de Datos.

Teniendo en cuenta la segunda iteración, se lleva a cabo un análisis de riesgo (AR2), donde se lleva a cabo el estudio y previsión de futuros riesgos que puedan llegar a ocurrir teniendo en cuenta lo descrito, modelado y desarrollado hasta el momento. En la tabla 4.2. se lleva a cabo la demostración de la relación entre los Procesos descritos en esta segunda vuelta del espiral propuesto y los riesgos identificados en [Britos, 2008].

Proceso del Espiral	Riesgos – Análisis Riesgo 2
Evaluación de la Situación, Objetivos del Proyecto de Explotación de Información y Plan de Proyecto	El cliente puede no estar seguro de lo que el proyecto puede hacer y lograr
	El cliente no sabía las necesidades de información de la organización
	El cliente está en un grupo impredecible (no tiene tanto interés por el proyecto)

Tabla 4.2. Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR2

4.2.1.3. Tercer Vuelta

Luego de la segunda iteración, del Análisis de Riesgo 2 (AR2) y teniendo como referencia los procesos que corren en paralelo y a lo largo de todo el Modelo de Ciclo de Vida como son las Simulaciones, los Modelos y los Benchmarks, como así también los Subprocesos de Administración de Proyectos [Vanrell, 2012] se lleva a cabo la tercera iteración partiendo del Eje de Desarrollo, Adaptación y Pruebas.

Durante esta iteración se llevan a cabo tareas correspondientes a la Fase de **Entendimiento de los Datos** de la Metodología CRISP-DM y de los Procesos de Desarrollo de Proyectos del Modelo de Procesos [Vanrell, 2012] para Proyectos de Explotación de Información.

En primer lugar se lleva a cabo la Recolección de Datos primarios o iniciales, donde se accede a los datos listados en los Recursos del Proyecto y donde se incluye carga de datos, si es necesario para la comprensión de los datos. Como salida perteneciente a esta actividad se obtiene el Informe de Colección de Datos Inicial, donde quedan listados el conjunto de datos obtenido junto con sus

posiciones, los métodos usados para adquirirlos, algunos de los problemas encontrados y algunas de las resoluciones alcanzadas.

Posteriormente se realiza la actividad de Descripción de los Datos donde se examina las propiedades "gruesas" o "superficiales" de los datos e informe adquiridos en los resultados. La salida de esta actividad es un Informe de descripción de datos donde se describen los datos que han sido adquiridos, incluyendo el formato de los datos, la cantidad de datos (por ejemplo, el número de registros y campos en cada tabla), los identificadores de los campos, y cualquier otro rasgo superficial que ha sido descubierto.

En tercer lugar se realiza la Exploración de los Datos donde se dirigen interrogantes de minería de datos usando preguntas, visualización, y técnicas de reporte incluyendo la distribución de atributos claves relacionados entre pares o pequeños números de atributos, los resultados de simples agregaciones, las propiedades de las subpoblaciones significativas, y análisis estadísticos simples. Como salida a esta actividad se obtiene un Informe de Exploración de Datos donde se describen los resultados de esta tarea, incluyendo primeras conclusiones o hipótesis iniciales y su impacto sobre el resto del proyecto.

Por último se lleva a cabo una Verificación de la Calidad de los Datos, donde se examina la calidad de los datos, realizando preguntas como: ¿Los datos están completos? ¿Son correctos, o estos contienen errores y, si hay errores, que tan comunes son estos? ¿Hay valores omitidos en los datos? Si es así, ¿cómo se representan estos, donde ocurre esto, y que tan comunes son estos?.

Al finalizar las actividades de la tercera vuelta en el espiral posicionados en el eje de Desarrollo, Adaptación y Pruebas, se pasa al eje de Planeamiento de la siguiente fase llevando a cabo la Planificación de Preparación de los Datos.

Teniendo en cuenta la tercera iteración y situados en el eje de Evaluación de Alternativas e identificación y resolución de riesgos, se lleva a cabo un análisis de riesgo (AR3), donde se lleva a cabo el estudio y previsión de futuros riesgos que puedan llegar a ocurrir teniendo en cuenta lo descrito, modelado y desarrollado hasta el momento. En la tabla 4.3. se lleva a cabo la demostración de la relación entre los Procesos descritos en esta segunda vuelta del espiral propuesto y los riesgos identificados en [Britos, 2008].

Proceso del Espiral	Riesgos – Análisis Riesgo 3
Recolección de Datos, Descripción de Datos, Exploración de Datos y Verificación de calidad de datos	Los datos identificados por los requisitos no eran los correctos

Tabla 4.3. Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR3

4.2.1.4. Cuarta Vuelta

En la cuarta vuelta del espiral propuesto se llevan a cabo las tareas pertenecientes a la fase de **Preparación de los Datos** de la metodología CRISP-DM.

En primer lugar se realiza la Selección de Datos, donde se decide que datos serán usados para el análisis. Los criterios incluyen la importancia a los objetivos de la minería de datos, la calidad, y las restricciones técnicas como límites sobre el volumen de datos o los tipos de datos. Como salida correspondiente a dicha tarea se obtiene un razonamiento para inclusión o exclusión de datos y los motivos de esa determinación.

Posteriormente se lleva a cabo la Limpieza de Datos, que implica la selección de los subconjuntos de datos limpios, la inserción de datos por defectos adecuados, o técnicas más ambiciosas tales como la estimación de datos faltantes mediante modelado. La salida asociada a esta tarea es un Informe de la Limpieza de Datos, donde se describe que decisiones y acciones fueron tomadas para dirigir los problemas de calidad de datos informados.

A continuación se construyen las operaciones de preparación de datos tales como la producción de atributos derivados o el ingreso de nuevos registros, o la transformación de valores para atributos existentes en la actividad Construcción de Datos. Las salidas correspondientes a esta actividad son Atributos derivados (atributos nuevos que son contruidos de uno o más atributos existentes en el mismo registro) y los Registros generados (creación de registros completamente nuevos).

Luego, en la actividad Integración de Datos, se definen los métodos por el cual la información es combinada de múltiples tablas o registros para crear nuevos registros o valores. Esta actividad tiene como salida la Combinación de Datos, en la cual se realiza la unión simultánea de dos o más tablas que tienen información diferente sobre el mismo objeto.

En cuarto lugar, se lleva a cabo el Formateo de Datos, en el cual se realizan modificaciones principalmente sintácticas hechas a los datos que no cambian su significado, pero podría ser requerido por la herramienta de modelado. A partir de dicha actividad se obtienen los Datos formateados.

Por último y para finalizar las actividades pertenecientes a la fase de Preparación de los Datos, se lleva a cabo la Validación de los mismos para determinar si los mismos son realmente correctos y servirán para el desarrollo del Proyecto de Explotación de Información.

Pasando al siguiente eje, se realiza la Planificación del Modelo y se lleva a cabo un análisis de riesgo (AR4), donde se lleva a cabo el estudio y previsión de futuros riesgos que puedan llegar a ocurrir teniendo en cuenta lo descrito, modelado y desarrollado hasta el momento. En la tabla 4.4. se lleva a cabo la demostración de la relación entre los Procesos descritos en esta segunda vuelta del espiral propuesto y los riesgos identificados en [Britos, 2008].

Proceso del Espiral	Riesgos – Análisis Riesgo 4
Selección de Datos, Limpieza de Datos, Construcción de Datos, Integración de Datos y Formateo de Datos.	Los datos identificados por los requisitos no eran los correctos

Tabla 4.4. Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR4

Luego del Análisis de Riesgo 4 (AR4) se lleva a cabo como en cada iteración del espiral simulaciones, modelos y benchmarks como así también se realizan los Subprocesos de Administración de Proyectos descriptos en la sección 2 (Vanrell, 2012).

4.2.1.5. Quinta Vuelta e iteraciones

A partir de aquí comienza las actividades pertenecientes a la fase de **Modelado** correspondiente a la metodología CRISP-DM y de los Procesos de Desarrollo del Modelo de Procesos [Vanrell, 2012] para Proyectos de Explotación de Información.. En esta fase se producen iteraciones, ya que se construirán n cantidad de modelos y prototipos hasta alcanzar el modelo que más se adecue a las pretensiones y requerimientos que el cliente ha brindado.

En primer lugar se lleva a cabo la Selección de Técnica de Modelado 1, donde se elige la técnica de modelado específico, por ejemplo, un árbol decisión construido con C4.5. Como salida a esta actividad se obtiene la Técnica de Modelado seleccionada (Documento la técnica de modelado real que está por ser usado) y Presunciones del Modelado (Muchas técnicas de modelado hacen presunciones específicas sobre los datos -por ejemplo, que todos los atributos tengan distribuciones uniformes, no encontrar valores no permitidos, el atributo de clase debe ser simbólico, etc).

A continuación de la Selección de la Técnica de Modelado 1 se realiza la Generación de Prueba de Diseño 1, donde se genera un procedimiento o el mecanismo para probar la calidad y validez del modelo. A partir de dicha actividad se genera la Prueba de Diseño, donde se describe el plan intencionado para el entrenamiento, la prueba, y la evaluación de los modelos.

Posteriormente, se realiza la primer Construcción del Modelo, donde se ejecuta la herramienta de modelado sobre el conjunto de datos preparados para crear uno o más modelos. Al finalizar dicha actividad se obtienen los parámetros de ajustes (Lista de los parámetros y sus valores escogidos, también con el razonamiento para elegir los parámetros de ajustes), el/los modelo/s en sí (modelos reales producidos por la herramienta de modelado, no un informe) y la descripción del/os modelo/s presentado/s (descripción de los modelos obtenidos e informe sobre la interpretación de los modelos y documentar cualquier dificultad encontrada con sus significados).

Al finalizar las 3 actividades mencionadas anteriormente, se lleva a cabo la primer Evaluación del Modelo (Evaluación del Modelo 1) donde se interpreta los modelos según el conocimiento de dominio, los criterios de éxitos de minería de datos, y el diseño de prueba deseado, se juzga el éxito

de la aplicación del modelado y se descubre técnicas mas adecuadas. Se evalúa los modelos según los criterios de evaluación. La salida es la Evaluación del/os Modelo/s en sí donde se resumen los resultados de esta tarea, se enlista las calidades de los modelos generados y se clasifica su calidad en relación con cada otro; y se obtienen los parámetros de ajustes revisados, donde se revisan los parámetros de ajuste para la siguiente corrida en la tarea de Construcción del Modelo.

Al finalizar este cuarto de vuelta, se realiza la Planificación de Evaluación y Despliegue 1 y en el eje de Evaluación de alternativas e identificación y resolución de Riesgos se lleva a cabo el Análisis de Riesgo numero 5 (AR5). En la tabla 4.5. se lleva a cabo la demostración de la relación entre los Procesos descritos en esta segunda vuelta del espiral propuesto y los riesgos identificados en [Britos, 2008].

Proceso del Espiral	Riesgos – Análisis Riesgo 5
Selección de Técnica de Modelado, Generación de Prueba de Diseño y Construcción del Modelo	Los modelos definidos pueden ser distintos a los que el cliente esperó.
	Cuando se está en fase de modelado y se detectan problemas en los datos, es necesario redefinir los requerimientos.
	Requisitos malentendidos del proyecto resultaron en la selección de la herramienta de modelado equivocada.

Tabla 4.5. Relación entre Proceso de Espiral y Riesgo que pueden aparecer en el AR5

Luego del AR5, se realiza el Prototipo 1, donde se lleva a cabo la demostración del Modelo construido y se analiza si cumple o no con lo que el cliente solicito.

Dependiendo de la satisfacción o cumplimiento del Modelo, se realizan 2 vueltas más al espiral, con las mismas actividades descritas anteriormente, hasta que finalmente el Modelo construido y desarrollado sea el adecuado para cumplir con los Objetivos planteados en un primer momento en el Proyecto.

Esas vueltas en blanco son previas a la conformación del Modelo final y aprobado y cada una de ellas es acompañada por un Análisis de Riesgo y la construcción de un Prototipo. Además, los subprocesos de Administración de Proyectos [Vanrell, 2012] acompañan estas tareas en paralelo.

En la anteúltima vuelta del espiral se lleva a cabo la ultima construcción del Modelo (Construcción Modelo n), donde n es la variable que indica cuantas veces se han llevado a cabo las actividades de la fase de Modelado hasta finalmente cumplir con los establecido.

Finalmente se pasa al eje de Planeamiento de la Siguiete fase para realizar la Planificación y Despliegue final (n), se pasa al siguiente eje realizando el Análisis de Riesgo n+4 (ya que previo al Prototipo 1 y AR5 se han llevado a cabo 4 análisis de Riesgo) y se construye el Prototipo final o Prototipo Operacional, que muestra el Modelo final o definitivo.

4.2.1.6. Última Vuelta

En la última vuelta del espiral propuesto se llevan a cabo las actividades de las fases de **Evaluación** y **Despliegue** de la metodología CRISP-DM y de los Procesos de Desarrollo del Modelo de Procesos [Vanrell, 2012] para Proyectos de Explotación de Información.

En primer lugar se Evalúan los Resultados, donde se analiza el grado al que el modelo responde a los objetivos de negocio y procura determinar si hay alguna decisión de negocio por el que este modelo es deficiente. Además verifica otros resultados generados por la minería de datos. La salida a esta actividad es la Evaluación de los Resultados de la minería de datos en lo que concierne a criterios del éxito del negocio, y por otro lado, los Modelos finalmente aprobados.

Por otro lado, en segundo lugar, se realiza la Revisión del Proceso, donde los modelos resultantes pasan a ser satisfactorios y a satisfacer las necesidades de negocio y se lleva a cabo una revisión más cuidadosa de los compromisos de la minería de datos para determinar si hay cualquier factor importante o tarea que de algún modo ha sido pasada por alto. Como salida a esta actividad se tiene la Revisión del Proceso propiamente dicha, donde se resume la revisión de proceso y destaca las actividades que han sido omitidas y/o aquellas que deberían ser repetidas.

Luego de la Revisión del Proceso, se Determinan los Próximos Pasos, donde según los resultados de la evaluación y la revisión de proceso, el equipo de proyecto decide como proceder. En esta instancia el equipo decide si hay que terminar este proyecto y tomar medidas sobre el desarrollo si es apropiado, tanto iniciar más iteraciones, o comenzar nuevos proyectos de minería de datos. Las salidas a esta actividad es una Lista de Posibles Acciones y Decisión (se describe la decisión en cuanto a cómo proceder, junto con el razonamiento).

Pasando a las actividades de Despliegue, en primer lugar se lleva a cabo el Desarrollo del Plan, donde de acuerdo al desarrollo de los resultados de minería de datos en el negocio, esta tarea toma los resultados de la evaluación y determina una estrategia para el desarrollo. La salida de esta actividad es el Plan propiamente dicho.

Luego del Plan descrito anteriormente se realiza un Plan de Supervisión y Mantenimiento, ya que son cuestiones importantes si los resultados de minería de datos son parte del negocio cotidiano y de su ambiente. Al finalizar esta tarea se obtiene el Plan propiamente dicho que resume la estrategia de supervisión y mantenimiento incluyendo los pasos necesarios y como realizarlos.

Posteriormente se pasa a la actividad de Producción de Informe Final, donde el líder del proyecto y su equipo sobrescribe un informe final. Según el plan de desarrollo, este informe puede ser sólo un resumen del proyecto y sus experiencias o esto puede ser una presentación final y comprensiva de los resultados de minería de datos.

La salida a esta tarea es el Informe definitivo y la Presentación final, donde puede haber una reunión en la conclusión del proyecto en el que los resultados son presentados verbalmente al cliente.

Por último, se lleva a cabo una Revisión General del Proyecto, donde se evalúa lo que fue correcto y lo que se equivocó, lo que fue bien hecho y lo que necesita para ser mejorado. La salida correspondiente es la Documentación de la experiencia de realización del Proyecto, donde se resumen las experiencias importantes ganadas durante el proyecto.

4.2.1.7. División por Fase del Modelo de Procesos

Luego de analizar cada vuelta del Modelo de Ciclo de Vida propuesto anteriormente, en este trabajo se cree necesario ilustrar al Modelo dividido por las fases del Modelo de Procesos para Proyectos de Explotación de Información.

En la imagen 4.2. se puede observar el Modelo de Ciclo de Vida anteriormente propuesto para la solución de los problemas descritos en la sección 3 dividido por las fases del Modelo de Procesos seleccionado en este trabajo, el cual posee un valor agregado a la metodología CRISP-DM, agregándole los subprocesos de Administración del Proyecto.

Estos subprocesos de Administración abarcan todo el periodo de Desarrollo, ya que como se ha dicho en la sección 2 tienen la función de recolectar información que le sea útil a los procesos de Desarrollo para aumentar la calidad de los mismos permitiendo realizar ajustes y mantener un estándar en la realización de distintos proyectos llevando a cabo la conversión de procesos caóticos y aislados en procesos repetitivos y controlados.

Como se puede observar, en las primeras dos vueltas del Espiral se llevan a cabo actividades correspondientes a la Fase Entendimiento del Negocio. Esta es la fase que más vueltas posee dado que se considero de suma importancia desarrollar un fuerte análisis de riesgos para que tanto los objetivos del negocio como del data mining estén bien comprendidos y el Plan de Proyecto contemple, y cumpla, con lo que el cliente realmente necesita, sin malos entendidos ni léxicos dispares.

Luego de las 2 vueltas por la fase de Entendimiento del Negocio el Modelo de Ciclo de Vida en Espiral propuesto posee una vuelta por la fase de Entendimiento de los Datos y otra por la fase de Preparación de los Datos.

Posteriormente a esas primeras cuatro vueltas, se llega a la de Modelado, donde se producen las iteraciones dado que si el Modelo propuesto y construido no cumple con los requisitos planteados por el cliente se vuelven a ejecutar las actividades pertenecientes a esa vuelta.

Una vez logrado y alcanzado el Modelo deseado, se pasa a la última vuelta donde se encuentran las actividades tanto de la fase de Evaluación como de Despliegue.

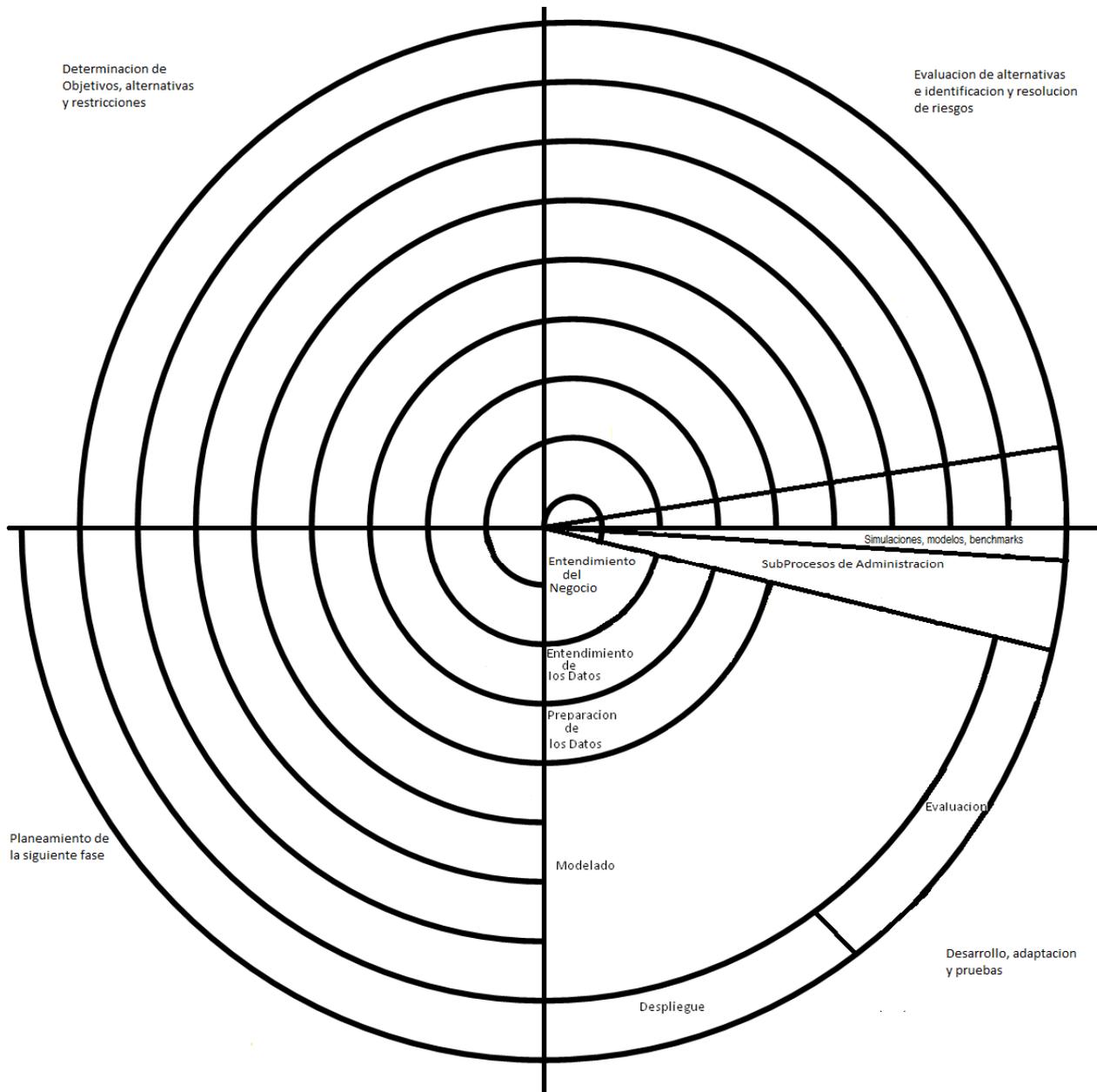


Figura 4.2. Separación del Modelo de Ciclo de Vida por procesos del Modelo de Procesos

4.2.2. Mapa de Actividades Propuesto

Considerando el Modelo de Ciclo de Vida propuesto en la subsección anterior, y teniendo como eje el Modelo de Procesos desarrollado por Vanrell [2012] para Proyectos de Explotación de Información, este trabajo lleva a cabo la construcción de un Mapa de Actividades teniendo como ejes los Modelos mencionados. En primer lugar se lleva a cabo un breve análisis del Mapa completo

(subsección 4.2.2.1.) para luego analizar en detalle las actividades de cada vuelta del Modelo de Ciclo de Vida (subsecciones 4.2.2.2. a 4.2.2.7.) cruzadas con los procesos y subprocesos del Modelo de Vanrell [2012]

4.2.2.1. Mapa de Actividades Completo

Tal como se mencionó anteriormente se presentan en las figuras 4.3.a) y 4.3. b) el Mapa de Actividades construido y propuesto, teniendo como ejes en primer lugar el Modelo de Ciclo de Vida propuesto que tuvo como base el Modelo de Ciclo de Vida en Espiral que propuso Boehm y las fases del Modelo de Procesos desarrollado por Vanrell para Proyectos de Explotación de Información, agregándole procesos de Administración de Proyectos los cuales la metodología existente, CRISP-DM, no poseía, y en segundo lugar el Modelo de Procesos de Vanrell propiamente dicho.

Como se puede observar, se dividen las actividades de los Procesos de Administración de Proyectos y los Procesos de Desarrollo de Proyectos, tal como describe Vanrell en su trabajo [Vanrell, 2012]. Por otro lado, cada columna es una actividad que se lleva a cabo en el Modelo de Ciclo de Vida propuesto en este trabajo, marcando la división por vuelta.

Las actividades de Administración de Proyectos no están ocupados del desarrollo propiamente dicho del Proyectos de Explotación de Información, sino que son procesos en paralelo y que, como se ha mencionado en el capítulo 2, tienen la función de recolectar información que le sea útil a los procesos de Desarrollo para aumentar la calidad de los mismos permitiendo realizar ajustes y mantener un estándar en la realización de distintos proyectos llevando a cabo la conversión de procesos caóticos y aislados en procesos repetitivos y controlados [Vanrell, 2012]. Por ello mismo se puede ver que las actividades de Administración de Proyectos abarcan todo el Ciclo de Vida, soportando y manteniendo las actividades de Desarrollo de Proyectos del Modelo de Vanrell.

Para aclarar y entender el Mapa de Actividades propuesto usaremos la siguiente convención:

- Fases del Modelo de Ciclo de Vida (Negrita y Subrayado)
- División Procesos de Administración y Desarrollo del Proyecto (Negrita)
- SubProcesos (Subrayado)
- Tarea (Itálica)

4.2.2.2. Mapa de Actividades – Primera Vuelta del Espiral

Como se ha mencionado en la introducción de la subsección 4.2.2., se describe el Mapa de Actividades por fases correspondiente a cada vuelta del Espiral del Modelo de Ciclo de Vida construido y presentado anteriormente. En cada una de estas subsecciones se analiza y justifica que actividades se realizan en cada fase y cuáles no.

ACTIVIDADES DE LOS PROCESOS	O.N.	E.Sit.	O.P.	P.P.	Val.O.	R.D.	D.D.	E.D.	Ver.CD	S.D.	L.D.	C.D.	I.D.	F.D.	Val.D.	S.T.M.	G.P.D.	C.M.	E.M.	E.R.	R.P.	D.P.P.	D.P.	P.S.M.	I.P.	R.Py.
Proceso de Administración de Proyectos																										
Planificación /Entendimiento del negocio																										
- Entendimiento del negocio		■	■	■																						
- Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento				■																						
- Definir el protocolo de entrega con el cliente				■																						
- 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 equipo de trabajo				■	■																					
- Establecer el calendario de actividades				■	■																					
- Calcular el costo estimado del proyecto				■	■																					
- Evaluación de la situación		■		■	■	■	■	■	■	■	■	■	■	■	■	■										
- Producir un Plan de Proyecto				■	■																					
- Producir un Plan de Desarrollo				■	■							■	■	■	■	■										
- Formalizar el inicio de un nuevo ciclo del proyecto				■	■										■											
Realización																										
- Acordar las tareas con el equipo de trabajo				■	■											■	■	■								
- Acordar la distribución de información				■	■												■	■								
- Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario				■	■													■	■							
- Revisar cumplimiento del plan de adquisiciones y capacitación				■	■														■							
- 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																										
- 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																										
- Formalizar la terminación del proyecto o ciclo																							■	■	■	■
- Llevar a cabo el cierre del contrato con subcontratistas																							■	■	■	■
- Generar el reporte de mediciones y sugerencias de mejora																							■	■	■	■
- Planear la entrega																							■	■	■	■

Figura 4.3 a) Mapa de Actividades Propuesto para Administración de Proyectos

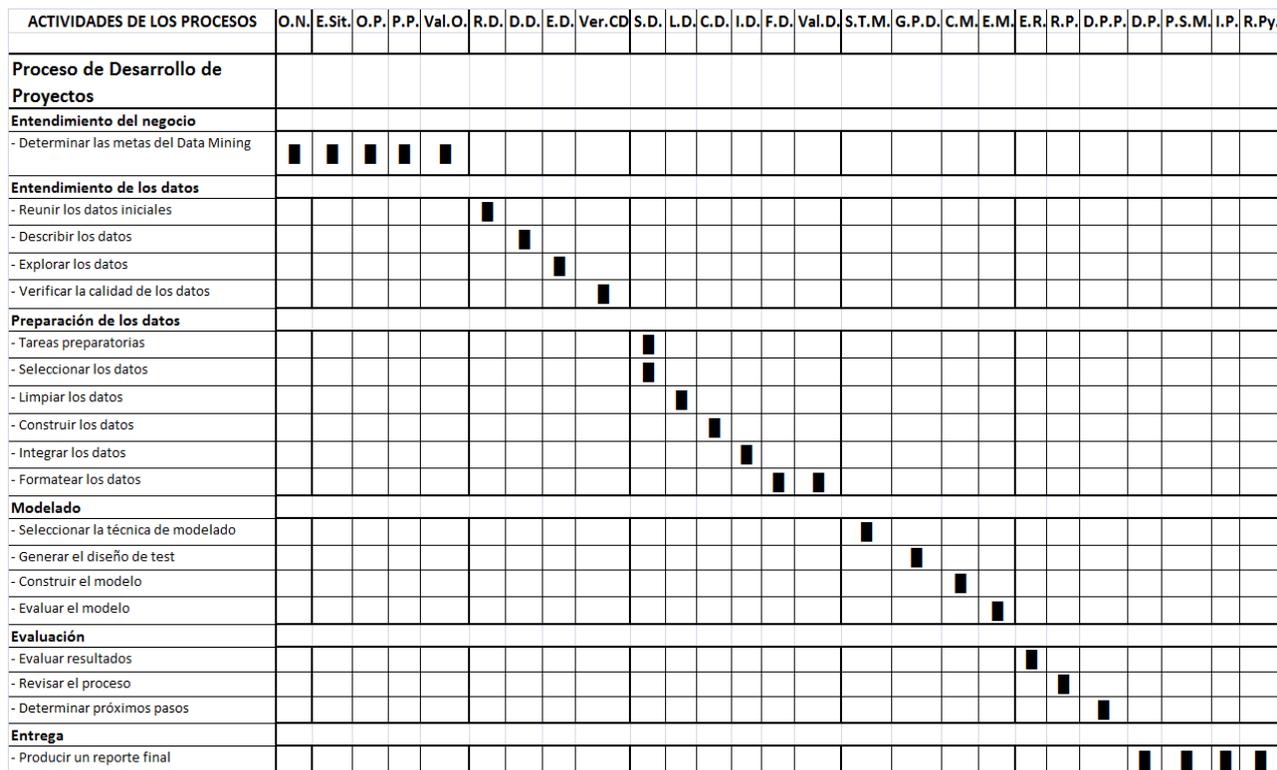


Figura 4.3 b) Mapa de Actividades Propuesto para Desarrollo de Proyectos

A continuación, en la tabla 4.6., se presentan las descripciones de los códigos de las fases del Modelo de Ciclo de Vida propuesto, utilizados en el Mapa de Actividades presentado

Código	Descripción
O.N.	Objetivos del Negocio
E.Sit.	Evaluación de la Situación
O.P.	Objetivos del Proyecto de E.I.
P.P.	Plan De Proyecto
Val.O.	Validación de Objetivos
R.D.	Recolección de Datos
D.D.	Descripción de Datos
E.D.	Exploración de Datos
Ver.CD.	Verificación Calidad de Datos
S.D.	Selección de Datos
L.D.	Limpieza de Datos
C.D.	Construcción de Datos
I.D.	Integración de Datos
F.D.	Formateo de Datos
Val.D.	Validación Datos
S.T.M.	Selección Técnica Modelado
G.P.D.	Generación Prueba de Diseño
C.M.	Construcción Modelo
E.M.	Evaluación Modelo
E.R.	Evaluación Resultados
R.P.	Revisión Proceso
D.P.P.	Determinación Próximos Pasos
D.P.	Desarrollo Plan
P.S.M.	Plan Supervisión y Mantenimiento
I.P.	Informe Producto
R.Py.	Revisión Proyecto

Tabla 4.6. Descripción de los Códigos de las Fases utilizados en las figuras 4.3.

En la imagen 4.4. se puede observar el Mapa de Actividades desarrollado pero teniendo en cuenta solo la primera vuelta del espiral del Modelo de Ciclo de Vida construido.

En la primera vuelta del espiral la única fase que se lleva a cabo es la de **Objetivos del Negocio**, donde se describen los objetivos primarios del cliente, desde una perspectiva del negocio. Teniendo en cuenta los procesos y tareas que se definen en el Modelo de Procesos para Proyectos de Explotación de Información propuesto por Vanrell [2012] se puede decir que dentro de los **Procesos de Administración de Proyectos** se encuentran la Planificación/Entendimiento del Negocio, la Realización, la Evaluación y Control y el Cierre/Entrega.

En primer lugar, dentro de la Planificación /Entendimiento del Negocio, la única actividad que se realiza en la Fase de Objetivos del Negocio es la de *Entendimiento del Negocio*. Esto se debe a que la salida correspondiente a esta tarea es la Base de Conocimiento del Negocio, cuyas actividades previas son la de por medio de organigramas, redes de expertos e informes relacionados con el negocio, desarrollar gráficos organizacionales, identificar personas claves, sponsor interno, si existe comité de gobierno, las unidades del negocio que serán impactadas, etc. Por otro lado, otra salida correspondiente a esta tarea son los Objetivos del Negocio propiamente dicho, describiendo el problema en términos generales, evaluando el estado actual del proyecto, especificando los requerimientos y preguntas del negocio, determinando los beneficios esperados en términos de negocio. Por último, dentro de aquí se determinan los Criterios del Éxito del negocio y se identifican quienes determinan dichos criterios. Por otro lado, no se han marcado las actividades de Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento; Definir el protocolo de entrega con el cliente; 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 equipo de trabajo; Establecer el calendario de actividades; Calcular el costo estimado del proyecto; Evaluación de la situación; Producir un Plan de Proyecto; Producir un Plan de Desarrollo; y Formalizar el inicio de un nuevo ciclo del proyecto dado que se ha entendido que dichas actividades no corresponden al Negocio en sí, sino mas a los Objetivos del Proyecto de Explotación de Información en sí mismo.

En segundo lugar, dentro de la Realización, se ha definido que no se lleva a cabo ninguna actividad en la Fase de Objetivos del Negocio, ya que son actividades mas relacionadas con el Proyecto de Explotación en sí, su realización y con el seguimiento y monitoreo que se realiza durante el desarrollo del Proyecto.

En tercer lugar, dentro de Evaluación y Control, se ha determinado que la única actividad que se hará en esta Fase es la de *Generar el reporte de seguimiento del proyecto*, dado que dentro de la misma se evalúan si los objetivos del negocio y del uso del modelo puede cambiar a lo largo del

tiempo y se documenta lo más completamente posible el problema inicial que el modelo está intentando resolver. Por otro lado, las tareas de Evaluar el cumplimiento del plan de proyecto y plan de desarrollo; y Analizar y controlar los riesgos no se contemplan en esta Fase, dado que son más ligados al Plan del Proyecto de Explotación de Información en sí que del Negocio.

Por último, dentro del Cierre / Entrega, se definió que no se lleva ninguna tarea dentro de esta Fase, dado que estas actividades tienen más que ver con las últimas fases del Modelo de Ciclo de Vida, donde se lleva a cabo el Cierre y Entrega del Proyecto de Explotación de Información desarrollado.

Dentro de los **Procesos de Desarrollo de Proyecto** definido por Vanrell en su Modelo de Procesos, se encuentran Entendimiento del Negocio, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Entrega.

En primer lugar, dentro del subproceso Entendimiento del Negocio, se ha definido que su actividad *Determinar las metas del Data Mining* ha de llevarse en la Fase Objetivos del Negocio del Modelo de Ciclo de Vida, dado que se trasladan los interrogantes del negocio a metas del proyecto de Explotación de Información y se realiza informe indicando una traducción de los interrogantes a metas de negocio.

Por otro lado, dentro de los demás subprocesos correspondiente a los procesos de Desarrollo del Proyecto, se ha determinado que no se lleva a cabo ninguna tarea dentro de esta fase, dado que las mismas tienen más que ver con fases siguientes y más avanzadas del Modelo de Ciclo de Vida propuesto.

4.2.2.3. Mapa de Actividades – Segunda Vuelta del Espiral

En esta subsección se analiza y justifica que actividades se realizan en cada fase correspondientes a la segunda vuelta del espiral y cuáles no.

En la imagen 4.5. se puede observar en detalle que actividades correspondientes al Modelo de Procesos se llevan a cabo y no en cada fase del Modelo de Ciclo de Vida.

Esta segunda vuelta del Modelo en Espiral cuenta con las siguientes fases: Evaluación de la Situación, Objetivos del Proyecto de Explotación de Información, Plan del Proyecto y Validación de Objetivos.

Empezando a analizar en detalle a la fase de **Evaluación de la Situación**, se puede determinar que dentro del subproceso Planificación/Entendimiento del Negocio correspondiente a los Procesos de **Administración de Proyectos**, las únicas actividades que se llevan a cabo son la de *Entendimiento del Negocio* y *Evaluación de la Situación*, ya que se entiende que dentro de ellas tareas tales como evaluar el estado actual del proyecto, identificar necesidades y expectativas de los usuarios, identificar los recursos hardware disponible, establecer si los planes de mantenimiento de hardware entran en conflicto con la disponibilidad de hardware necesaria para el proyecto de Explotación de

Información, identificar las fuentes de datos, chequear las disponibilidades de herramientas y técnicas, identificar analistas de mercado, entre otras, resultan necesarias realizarlas en dicha frase.

ACTIVIDADES DE LOS PROCESOS	Objetivos del Negocio
Proceso de Administración de Proyectos	
Planificación /Entendimiento del negocio	
- Entendimiento del negocio	■
- Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento	
- Definir el protocolo de entrega con el cliente	
- 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 equipo de trabajo	
- Establecer el calendario de actividades	
- Calcular el costo estimado del proyecto	
- Evaluación de la situación	
- Producir un Plan de Proyecto	
- Producir un Plan de Desarrollo	
- Formalizar el inicio de un nuevo ciclo del proyecto	
Realización	
- Acordar las tareas con el equipo de trabajo	
- Acordar la distribución de información	
- Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario	
- Revisar cumplimiento del plan de adquisiciones y capacitación	
- 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	
- 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	
- Formalizar la terminación del proyecto o ciclo	
- Llevar a cabo el cierre del contrato con subcontratistas	
- Generar el reporte de mediciones y sugerencias de mejora	
- Planear la entrega	
Proceso de Desarrollo de Proyectos	
Entendimiento del negocio	
- Determinar las metas del Data Mining	■
Entendimiento de los datos	
- Reunir los datos iniciales	
- Describir los datos	
- Explorar los datos	
- Verificar la calidad de los datos	
Preparación de los datos	
- Tareas preparatorias	
- Seleccionar los datos	
- Limpiar los datos	
- Construir los datos	
- Integrar los datos	
- Formatear los datos	
Modelado	
- Seleccionar la técnica de modelado	
- Generar el diseño de test	
- Construir el modelo	
- Evaluar el modelo	
Evaluación	
- Evaluar resultados	
- Revisar el proceso	
- Determinar próximos pasos	
Entrega	
- Producir un reporte final	

Figura 4.4. Mapa de Actividades Propuesto para la primera vuelta del espiral.

Las otras actividades correspondientes a este SubProceso, se entiende que no corresponde a esta fase analizada ya que se cree que se relacionan mas con fases próximas de esta misma vuelta del Espiral (como por ejemplo producir plan de proyecto y formalizar el inicio de un nuevo ciclo).

Siguiendo con el análisis correspondiente a la fase de Evaluación de la Situación, se dictamina que las actividades pertenecientes tanto al SubProceso Realización, Evaluación y Control, y Cierre /Entrega, no se llevan a cabo dentro de esta fase puesto que la misma es prematura y dichas actividades corresponden cuando el Proyecto esta mas avanzado y se encuentra en etapas posteriores.

Dentro de los **Procesos de Desarrollo de Proyectos**, en el único Proceso que se llevan a cabo actividades en esta Fase es en el de Entendimiento del Negocio. Dentro de este SubProceso se lleva a cabo la tarea de *Determinar las metas del Proyecto de Explotación de Información*. Tareas de SubProcesos siguientes tales como Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Entrega no se llevan a cabo en esta Fase del Modelo de Ciclo de Vida ya que es un fase inicial y dichos procesos corresponden a tareas posteriores.

Siguiendo con el análisis con la segunda vuelta del Modelo de Ciclo de Vida en Espiral propuesto, la fase que se analiza a continuación es la de **Objetivos del Proyecto de Explotación de Información**.

Dentro de los Procesos de **Administración de Proyectos**, las únicas tareas que se llevan a cabo en el SubProceso Planificación/Entendimiento del Negocio son las de *Entendimiento de Negocio y Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento*.

Se ha definido esto ya que tareas tales como evaluar la descripción del proyecto, evaluar el proceso de desarrollo y mantenimiento, describir el problema en términos generales, clasificar prerrequisitos del proyecto, especificar los beneficios esperados y definir el proceso específico a utilizar, son incluidas dentro de esta fase.

Por otro lado, las tareas correspondiente al SubProceso de Realización no se realiza ninguna salvo las de *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*, ya que recibir solicitudes de cambio del cliente y las Realizaciones de reuniones con el cliente pueden afectar y es importante tener en cuenta los Objetivos establecidos para el Proyecto de Explotación de Información.

Dentro del SubProceso de Evaluación y Control, la actividad que se lleva a cabo es la de *generar el reporte de seguimiento del cliente*, ya que se analiza que cosas pueden cambiar a futuro. Por otro lado, dentro del SubProceso de Cierre/Entrega, no se lleva a cabo ninguna actividad ya que se entiende que corresponden a fases posteriores.

Siguiendo con el análisis de esta fase, dentro de los Procesos de **Desarrollo de Proyecto**, el único SubProceso que se relaciona con esta fase es la de Entendimiento del Negocio. Es así ya que determinar las metas del proyecto de Explotación de Información especifica los criterios de valoración de los modelos, determina los benchmarks para la evaluación de criterios y especifica criterios a valoración de criterios subjetivos, entre otras tareas.

Las otras actividades que corresponden a los demás SubProcesos de Desarrollo de Proyectos, se entiende que no corresponden a esta fase ya que son actividades que se deben desarrollar fases posteriores.

La siguiente fase a analizar es la de **Plan del Proyecto**. Dentro de los **Procesos de Administración del Proyecto**, en esta fase, salvo en la tarea de Entendimiento del Negocio, se llevan a cabo todas las tareas tanto del SubProceso de Planificación /Entendimiento del Negocio como de Realización. Se lleva a cabo esta asunción ya que todas esas actividades influyen en el Plan de Proyecto desarrollado, tales como la *generación de una base de conocimiento del negocio, la definición de los objetivos del negocio, definición de criterios del éxito, desarrollo de Plan de Entrega, Plan de Desarrollo, Plan de Adquisiciones, armado de Calendario de actividades, estimación de costos, generación de inventario de recursos, riesgos y contingencias, descripción de la terminología empleada, lista de técnicas, reporte de seguimiento y monitoreo*.

Por otro lado, dentro de la Evaluación y Control, la actividad que se lleva a cabo es la de *generar el reporte de seguimiento del proyecto*, ya que afecta al Plan de Proyecto. Las actividades de cierre/entrega no se realizan en esta fase ya que no corresponde a esta altura.

Dentro de los **Proceso de Desarrollo del Proyecto**, en el único SubProceso donde se lleva a cabo una actividad en esta fase es en el de Entendimiento del Negocio, debido a que las metas del proyecto de Explotación de Información y los Criterios del Éxito del Proyecto de E.I. quedan registrados en la conformación del Plan de Proyecto.

Las demás actividades correspondientes a los demás SubProcesos de Desarrollo del Proyecto no se llevan a cabo en esta fase ya que son actividades posteriores a las configuraciones del Plan de Proyecto.

La última fase dentro de la segunda vuelta que se lleva a cabo en el Modelo de Ciclo de Vida en Espiral propuesto es la de **Validación de Objetivos**.

Dentro de los **Procesos de Administración del Proyecto**, en esta fase al igual que la fase anterior, salvo en la tarea de Entendimiento del Negocio y la de definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento, se llevan a cabo todas las tareas tanto del SubProceso de Planificación /Entendimiento del Negocio como de Realización.

ACTIVIDADES DE LOS PROCESOS	Evaluacion de la Situacion	Objetivos del P. Proyecto de E.I.	Plan del Proyecto	Validacion de Objetivos
Proceso de Administración de Proyectos				
Planificación /Entendimiento del negocio				
- Entendimiento del negocio	■	■		
- Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento		■		
- Definir el protocolo de entrega con el cliente			■	
- 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 equipo de trabajo			■	■
- Establecer el calendario de actividades			■	■
- Calcular el costo estimado del proyecto			■	■
- Evaluación de la situación	■		■	■
- Producir un Plan de Proyecto			■	■
- Producir un Plan de Desarrollo			■	■
- Formalizar el inicio de un nuevo ciclo del proyecto			■	■
Realización				
- Acordar las tareas con el equipo de trabajo			■	■
- Acordar la distribución de información			■	■
- Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario			■	■
- Revisar cumplimiento del plan de adquisiciones y capacitación			■	■
- 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				
- 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				
- Formalizar la terminación del proyecto o ciclo				
- Llevar a cabo el cierre del contrato con subcontratistas				
- Generar el reporte de mediciones y sugerencias de mejora				
- Planear la entrega				
Proceso de Desarrollo de Proyectos				
Entendimiento del negocio				
- Determinar las metas del Data Mining	■	■	■	■
Entendimiento de los datos				
- Reunir los datos iniciales				
- Describir los datos				
- Explorar los datos				
- Verificar la calidad de los datos				
Preparación de los datos				
- Tareas preparatorias				
- Seleccionar los datos				
- Limpiar los datos				
- Construir los datos				
- Integrar los datos				
- Formatear los datos				
Modelado				
- Seleccionar la técnica de modelado				
- Generar el diseño de test				
- Construir el modelo				
- Evaluar el modelo				
Evaluación				
- Evaluar resultados				
- Revisar el proceso				
- Determinar próximos pasos				
Entrega				
- Producir un reporte final				

Figura 4.5. Mapa de Actividades Propuesto para la segunda vuelta del espiral.

Se lleva a cabo esta asunción ya que todas esas actividades influyen en el Plan de Proyecto desarrollado, tales como la *generación de una base de conocimiento del negocio, la definición de los objetivos del negocio, definición de criterios del éxito, desarrollo de Plan de Entrega, Plan de Desarrollo, Plan de Adquisiciones, armado de Calendario de actividades, estimación de costos, generación de inventario de recursos, riesgos y contingencias, descripción de la terminología empleada, lista de técnicas, reporte de seguimiento y monitoreo.*

Dentro de la Evaluación y Control se lleva a cabo la tarea de *generar el reporte de seguimiento del cliente*, ya que se evalúan si los objetivos pueden cambiar a lo largo del Proyecto.

En los **Procesos de Desarrollo del Proyecto** la única actividad que se realiza en esta fase es la de *determinar las metas del Data Mining*, correspondiente al SubProceso de Entendimiento del Negocio. Las demás actividades de los restantes 5 SubProcesos no corresponden a esta fase ya que como se ha mencionado en otras fases, esas tareas son ligadas a etapas posteriores.

4.2.2.4. Mapa de Actividades – Tercera Vuelta del Espiral

Continuando con en el análisis del Mapa de Actividades para Proyectos de Explotación de Información construido, se pasa a la evaluación de la tercer vuelta correspondiente al Modelo de Ciclo de Vida en Espiral propuesto (figura 4.6.).

La primera fase de esta vuelta del espiral es la de **Recolección de Datos**. Teniendo en cuenta los **Procesos de Administración de Proyectos**, dentro de la Planificación/Entendimiento del Negocio, la única tarea que se lleva a cabo es la de *Evaluación de la Situación*. Esto es así ya que tareas como identificar las fuentes de datos, identificar los distintos tipos de fuentes de datos, identificar fuentes de conocimiento, chequear la disponibilidad de técnicas y herramientas, identificar los administradores de sistemas, bases de datos, entre otras, se vinculan estrechamente con esta Fase. Por otro lado, las actividades anteriores y posteriores de este SubProceso no se realizan en esta fase del Modelo ya que se entiende que deben hacerse en otras fases debido al tiempo de desarrollo del Proyecto. Las actividades pertenecientes a los SubProcesos de Realización, Evaluación y Control, y Cierre/Entrega, no se vinculan con esta fase puesto que son actividades que se llevan a cabo en fases posteriores y no se relacionan con la Recolección de Datos propiamente dicha.

Entrando en los **Procesos de Desarrollo de Proyectos**, se define que las actividades pertenecientes al SubProceso de Entendimiento del Negocio no se llevan a cabo en esta fase ya que son tareas propias de las 2 primeras vueltas del espiral. Sin embargo, la tarea de *Reunir los Datos Iniciales* del SubProceso de Entendimiento de los Datos sí se realiza ya que se identifica cual es la información necesaria, se chequea si la información está disponible, se especifica el criterio de selección, se seleccionan datos dentro de tablas o archivos, se evalúa como se debe adquirir datos faltantes, se

analiz si los datos contienen entradas de texto libre evaluar, si es necesario codificarla para el modelado o su es necesario agrupar entradas específicas, y se describe como se debe extraer los datos entre otras tareas.

Por otro lado, las demás actividades de este SubProceso y de los SubProcesos de Preparación de los Datos, Modelado, Evaluación y Entrega no se llevan a cabo ya que son tareas de fases siguientes.

La próxima fase a analizar es la de **Descripción de Datos**. Empezando por los **Procesos de Administración de Proyectos**, la única tarea que se lleva a cabo es la de *Evaluación de la Situación*, dentro del SubProceso de Planificación/Entendimiento del Negocio. Esto se define así ya que entre otras tareas se identifican las fuentes de datos, se describe el conocimiento (background), se listan las asunciones de la calidad de los datos, se chequean las restricciones generales en torno a los datos, se chequea la accesibilidad que los datos poseen, etc. Las actividades siguientes de este SubProceso no se asignan a esta fase ya que se entiende que tienen que ver más con el Negocio que con los Datos y su análisis en sí. Además, los demás SubProcesos no se relacionan con esta Fase ya que cubren más a otras que se llevan a cabo más adelante.

Considerando los **Procesos de Desarrollo de Proyecto**, las actividades del SubProceso de Entendimiento de Negocio no se relaciona con esta fase dado que son actividades de fases anteriores. En cambio, la actividad de *describir los datos* del SubProceso de Entendimiento de los datos sí, ya que se identifican los datos y métodos de captura, se acceden a las fuentes de datos, se usan análisis estadísticos, se chequea el volumen de datos, se chequea la disponibilidad y accesibilidad, se chequean los tipos de atributos, se analiza la correlación de atributos, entre otras tareas. Por otro lado, las demás actividades de este SubProceso y las actividades de Preparación de los Datos, Modelado, Evaluación y Entrega se realizan en fases posteriores, cuando el Proyecto este más avanzado en cuanto a su desarrollo.

Siguiendo con las fases correspondientes a la tercer vuelta, se analiza la **Exploración de Datos**. En primer lugar, dentro de los **Procesos de Administración de Proyectos**, se lleva a cabo en el SubProceso de Planificación /Entendimiento del Negocio la tarea de *Evaluación de la Situación*, tal como las dos fases anteriores de esta vuelta del Modelo de Ciclo de Vida en Espiral, ya que se chequean los derechos de acceso a las fuentes de datos, se chequean las técnicas de acceso a los datos, si el conocimiento relevante es accesible, etc. Por otro lado, las actividades de los SubProcesos siguientes no se realizan en esta fase ya que dan soporte y se relacionan con fases más avanzadas.

Teniendo en cuenta los **Procesos de Desarrollo de Proyectos**, el SubProceso que lleva a cabo una actividad en esta fase es la de Entendimiento del Negocio. Dentro del mismo, se realiza la *exploración de datos*, o sea analizar propiedades de atributos interesantes en detalle, identificar

características de sub-poblaciones, considerar y evaluar información y resultados en los reportes de descripción de datos, formular hipótesis e identificar acciones, transformar hipótesis en metas de Explotación de Información si es posible, clarificar las metas de Explotación de Información o hacerlas más precisas y realizar análisis básicos para verificar las hipótesis. Las actividades de los SubProcesos siguientes se realizan en fases más adelantadas por su avance en el desarrollo del Proyecto de E.I. La última fase a analizar en esta vuelta del Modelo de Ciclo de Vida en Espiral es la de **Verificación Calidad de Datos**. En primer lugar, teniendo en cuenta los **Procesos de Administración de Proyectos**, se lleva a cabo la tarea de *Evaluación de la Situación* del SubProceso de Planificación/Entendimiento del Negocio, ya que se listan las presunciones de calidad de los datos. Luego, dentro del SubProceso de Realización, se realiza la tarea de *recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo*. Se informa que esta actividad se realiza en esta fase dado que se generan un informe con los datos recolectados, se realizan los ajustes necesarios, se promueven las sugerencias de mejora en caso de ser adecuadas, se recolectan productos de trabajo, entre otras. Estas actividades influyen en la verificación de la calidad de datos. Las actividades de los demás SubProcesos correspondientes a la Administración de Proyecto corresponden a fases más avanzadas.

Por otro lado y finalizando con el análisis de la tercera vuelta del Modelo de Ciclo de Vida en Espiral, la actividad que se desarrolla en esta fase dentro de los **Procesos de Desarrollo de Proyectos**, es la de *verificar la calidad de datos* propiamente dicho, correspondiente al SubProceso de Entendimiento de los Datos. Dentro de esta actividad se incluyen tareas tales como identificar valores especiales y catalogar su significado, chequear claves, chequear cubrimientos, identificar atributos faltantes, chequear desviaciones, dar significado a la falta de datos, planear como lidiar con los ruidos, detectar tipos de ruidos, entre otros. Este tipo de actividad es propia de esta fase y de ninguna otra. Las demás actividades de los SubProcesos continuos corresponden a fases siguientes del Modelo de Ciclo de Vida propuesto.

4.2.2.5. Mapa de Actividades – Cuarta Vuelta del Espiral

En esta subsección se analiza y justifica que actividades se realizan en cada fase correspondientes a la cuarta vuelta del espiral y cuáles no.

En la imagen 4.7. se puede observar en detalle que actividades correspondientes al Modelo de Procesos se llevan a cabo y no en cada fase del Modelo de Ciclo de Vida.

Esta cuarta vuelta del Modelo en Espiral cuenta con las siguientes fases: Selección de Datos, Limpieza de Datos, Construcción de Datos, Integración de Datos, Formateo de Datos y Validación de Datos.

ACTIVIDADES DE LOS PROCESOS	Recoleccion de Datos	Descripción de Datos	Exploración de Datos	Verificación Calidad de Datos
Proceso de Administración de Proyectos				
Planificación /Entendimiento del negocio				
- Entendimiento del negocio				
- Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento				
- Definir el protocolo de entrega con el cliente				
- 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 equipo de trabajo				
- Establecer el calendario de actividades				
- Calcular el costo estimado del proyecto				
- Evaluación de la situación	■	■	■	■
- Producir un Plan de Proyecto				
- Producir un Plan de Desarrollo				
- Formalizar el inicio de un nuevo ciclo del proyecto				
Realización				
- Acordar las tareas con el equipo de trabajo				
- Acordar la distribución de información				
- Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario				
- Revisar cumplimiento del plan de adquisiciones y capacitación				
- 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				
- 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				
- Formalizar la terminación del proyecto o ciclo				
- Llevar a cabo el cierre del contrato con subcontratistas				
- Generar el reporte de mediciones y sugerencias de mejora				
- Planear la entrega				
Proceso de Desarrollo de Proyectos				
Entendimiento del negocio				
- Determinar las metas del Data Mining				
Entendimiento de los datos				
- Reunir los datos iniciales	■			
- Describir los datos		■		
- Explorar los datos			■	
- Verificar la calidad de los datos				■
Preparación de los datos				
- Tareas preparatorias				
- Seleccionar los datos				
- Limpiar los datos				
- Construir los datos				
- Integrar los datos				
- Formatear los datos				
Modelado				
- Seleccionar la técnica de modelado				
- Generar el diseño de test				
- Construir el modelo				
- Evaluar el modelo				
Evaluación				
- Evaluar resultados				
- Revisar el proceso				
- Determinar próximos pasos				
Entrega				
- Producir un reporte final				

Figura 4.6. Mapa de Actividades Propuesto para la tercera vuelta del espiral.

La primera fase en analizar en esta vuelta del espiral es la de **Selección de Datos**. Teniendo en cuenta los **Procesos de Administración de Proyectos**, en el único SubProceso donde se lleva a cabo una actividad es en la **de Planificación /Entendimiento del Negocio**. La actividad que se realiza en esta fase es la de *Evaluación de la Situación*, dado que registran las fuentes de datos, fuentes de conocimientos, se describen los datos, se identifican tipos de fuentes de conocimientos, entre otras. Las actividades de los demás SubProcesos no han de llevarse a cabo en dicha fase dado que no se relacionan directamente con la selección de datos propiamente dicha y además, dan soporte a fases posteriores en el desarrollo del proyecto.

Teniendo en cuenta los **Procesos de Desarrollo de Proyectos**, esta fase realiza las siguientes actividades pertenecientes al SubProceso de **Preparación de los Datos**: *Tareas Preparatorias y Selección de Datos*. Con relación a la primer actividad mencionada, se ha de llevar a cabo la reunión de datos producidos por la fase de preparación de datos y, con respecto a la selección de datos propiamente dicha, se recolectan datos adicionales apropiados, se realizan test de significancia, se reconsideran los criterios de selección de datos, se seleccionan diferentes subconjuntos de datos, se documentan razones de exclusión e inclusión, etc. Se ha definido estas actividades para esta fase ya que se relacionan con la selección de datos y no las demás ya que son propias de fases siguientes, como limpieza, construcción, integración y formateo.

En segundo lugar, analizando las actividades vinculadas con la fase de **Limpieza de Datos**, se define que tal como con la Selección de Datos, el único SubProceso de **Administración de Proyecto** que tiene actividades en esta fase es la de **Planeamiento /Entendimiento del Negocio**, ya que se realiza la *Evaluación de la Situación y la Producción de un Plan de Desarrollo*. En primer lugar la evaluación de la situación da soporte a la generación de limpieza de los datos obtenidos con las listas de asunciones y de riesgos producidas. En segundo lugar, en la Producción de un Plan de Desarrollo quedan registrados los pasos críticos y técnicas de limpieza que se llevan a cabo.

Por otro lado, teniendo en cuenta los **Procesos de Desarrollo de Proyectos**, se realiza la actividad de *limpiar los datos* pertenecientes al SubProceso de **Preparación de los Datos**, dado que se reconsidera como lidiar con los tipos de ruidos, se corrigen, remueven e ignoran los ruidos, se decide como lidiar con valores especiales y se reconsidera la selección de datos con base a experiencia de limpieza. Como se ha mencionado en fases anteriores, las actividades tanto de SubProceso de Modelado y Evaluación están relacionadas con fases mas avanzadas del espiral.

Siguiendo con el análisis de la cuarta vuelta del Modelo de Ciclo de Vida en Espiral propuesto en este capítulo, la siguiente fase en ser analizada es la de **Construcción de Datos**. Con respecto a los **Procesos de Administración de Proyectos**, se lleva a cabo dentro del SubProceso de **Preparación /Entendimiento del Negocio** la actividad de *Evaluación de la Situación y la de Producción de un*

Plan de Desarrollo. Ambas actividades dan soporte a nivel administrativo a la construcción de datos y sus técnicas de transformaciones.

Teniendo en cuenta los **Procesos de Desarrollo de Proyectos**, se lleva a cabo la actividad de *construir los datos*, dentro del SubProceso de Preparación de los Datos, chequeando los mecanismos de construcción disponibles, decidiendo si es mejor realizar la construcción dentro o fuera de la herramienta, reconsiderando los criterios de selección de datos, decidiendo si hay que normalizar atributos, decidiendo como pueden ser construidos o imputados atributos faltantes, realizando pasos de transformación o chequeando técnicas disponibles de generación de registros si son necesarios.

La siguiente fase de esta vuelta del espiral es la de **Integración de Datos**. Entrando en los **Procesos de Administración de Proyecto**, las actividades que se realizan tal como en las fases anteriores son las de *Evaluación de la Situación y la de Producción de un Plan de Desarrollo*, del SubProceso de Preparación/Entendimiento del Negocio. Tal como se menciona en la fase anterior, estas actividades dejan sentadas a nivel administrativo las técnicas de integración de los datos.

Las actividades posteriores a estos procesos corresponden a fases avanzadas del proyecto desarrollado.

Por otro lado, dentro de los **Procesos de Desarrollo de Proyectos**, específicamente en el SubProceso de Preparación de los Datos, se realiza la actividad de *integrar los datos* propiamente dicha. Esto es ya que se chequean si las facilidades de integración son capaces de integrar las fuentes de entrada como es requerido, se integran las fuentes y almacenan los resultados, y se reconsideran los criterios de selección de los datos basándose en las experiencias de integración.

La quinta fase en ser analizada de esta vuelta del Modelo en Espiral es la de **Formateo de Datos**. Teniendo en cuenta los **Procesos de Administración de Proyectos**, se realiza la actividad de *Evaluar la Situación y la de Producir un Plan de Desarrollo* del SubProceso de Planeamiento / Entendimiento del Negocio.

Dentro de los **Procesos de Desarrollo de Proyecto**, se realiza la actividad de *formatear los datos* que se realiza en el SubProceso de Preparación de los Datos. Las tareas que se llevan a cabo son las de reordenar los atributos si es requerido por las herramientas utilizadas, reordenar los registros si es necesario, reformatear valores para satisfacer los requerimientos de las herramientas utilizadas, entre otras.

La última fase de esta vuelta del espiral es la de **Validación de los Datos**. Con respecto a los **Procesos de Administración de Proyectos** se realizan las actividades de *evaluar la situación, producir un plan de proyecto, producir un plan de desarrollo y formalizar el inicio de un nuevo ciclo del proyecto*. A nivel gerencial y administrativo se da soporte por medio de estas actividades a

la validación de los datos obtenidos y seleccionados, quedando registrados que los datos son válidos y cumplen con los requerimientos del Proyecto. Por medio de la validación exitosa de los datos, se da inicio al ciclo correspondiente del proyecto y se le comunica a los interesados.

Por otro lado, dentro de los **Procesos de Desarrollo de Procesos**, la actividad que se vincula con esta fase es la de *formatear los datos*, dado que es la última actividad del SubProceso de Preparación de los Datos, y por medio de esta ya quedan los datos listos para ser usados en la construcción del modelo y finalmente se validan para corroborar si realmente cumplen con los requerimiento establecidos en el proyecto.

4.2.2.6. Mapa de Actividades – Quinta Vuelta del Espiral

En la quinta vuelta del Modelo de Ciclo de Vida en Espiral propuesto en este capítulo se encuentran las siguientes fases: Selección Técnica de Modelado, Generación Prueba de Diseño, Construcción Modelo y Evaluación del Modelo.

En la figura 4.8 se puede observar el Mapa de Actividades desarrollado enfocado en estas 4 fases mencionadas anteriormente.

Comenzando por la fase de **Selección Técnica de Modelado**, se puede observar que dentro de los **Procesos de Administración de Proyectos**, específicamente en el SubProceso de Realización, posee la actividad de *acordar las tareas con el equipo del trabajo*, puesto que se revisan las tareas a realizar, se evalúan los roles y se distribuyen las tareas. Por medio de estas actividades se brinda soporte a las actividades de desarrollo propias como la selección de la técnica a utilizar para desarrollar el modelo.

Teniendo en cuenta los **Procesos de Desarrollo de Proyectos** y dentro del SubProceso Modelado, se realiza la actividad de *seleccionar la técnica de modelado* propiamente dicha, mediante la cual se decide sobre la técnica apropiada para ejercitar teniendo en mente la herramienta seleccionada, se define cualquier asunción hecha por la técnica acerca de los datos, se comparan las asunciones con las hechas en el reporte de descripción de dato, y se asegura que las asunciones se mantienen y regresar a la fase de preparación de datos si es necesario.

En segundo lugar, siguiendo con la fase de **Generación Prueba de Diseño** y dentro de los **Procesos de Administración de Proyectos** se realizan las actividades de *acordar las tareas con el equipo del trabajo* y *acordar la distribución de información*, pertenecientes al SubProceso Realización. Por medio de estas tareas se distribuyen las tareas, se identifica la información requerida para cada tarea y cada plan de prueba del modelo a construir, y se distribuye la información a todo el equipo responsable del desarrollo del Proyecto. Las actividades siguientes de

este SubProceso y de los SubProcesos siguientes corresponden a fases mas avanzadas donde el Modelo ya fue desarrollado y lo que queda son tareas tales como evaluación y entrega al cliente.

ACTIVIDADES DE LOS PROCESOS	Selección de Datos	Limpieza de Datos	Construcción de Datos	Integración de Datos	Formateo de Datos	Validación de Datos
Proceso de Administración de Proyectos						
Planificación /Entendimiento del negocio						
- Entendimiento del negocio						
- Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento						
- Definir el protocolo de entrega con el cliente						
- 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 equipo de trabajo						
- Establecer el calendario de actividades						
- Calcular el costo estimado del proyecto						
- Evaluación de la situación	■	■	■	■	■	■
- Producir un Plan de Proyecto						■
- Producir un Plan de Desarrollo		■	■	■	■	■
- Formalizar el inicio de un nuevo ciclo del proyecto						■
Realización						
- Acordar las tareas con el equipo de trabajo						
- Acordar la distribución de información						
- Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario						
- Revisar cumplimiento del plan de adquisiciones y capacitación						
- 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						
- 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						
- Formalizar la terminación del proyecto o ciclo						
- Llevar a cabo el cierre del contrato con subcontratistas						
- Generar el reporte de mediciones y sugerencias de mejora						
- Planear la entrega						
Proceso de Desarrollo de Proyectos						
Entendimiento del negocio						
- Determinar las metas del Data Mining						
Entendimiento de los datos						
- Reunir los datos iniciales						
- Describir los datos						
- Explorar los datos						
- Verificar la calidad de los datos						
Preparación de los datos						
- Tareas preparatorias	■					
- Seleccionar los datos	■					
- Limpiar los datos		■				
- Construir los datos			■			
- Integrar los datos				■		
- Formatear los datos					■	■
Modelado						
- Seleccionar la técnica de modelado						
- Generar el diseño de test						
- Construir el modelo						
- Evaluar el modelo						
Evaluación						
- Evaluar resultados						
- Revisar el proceso						
- Determinar próximos pasos						
Entrega						
- Producir un reporte final						

Figura 4.7. Mapa de Actividades Propuesto para la cuarta vuelta del espiral.

Por otro lado, dentro de los **Procesos de Desarrollo de Proyectos**, se realiza la actividad de *generar el diseño del test*, en el SubProceso de Modelado, donde se chequean los diseños de test existentes para cada meta de E.I. de forma separada, se deciden los pasos necesarios y se preparan los datos requeridos para testear. Las siguientes actividades de estos Procesos corresponden a fases siguientes donde el Modelo ya está construido.

La tercera fase de esta vuelta del espiral es la de **Construcción del Modelo**. En los **Procesos de Administración de Proyectos** se llevan a cabo las actividades de *acordar las tareas con el equipo de trabajo, acordar la distribución de la información y revisar con el responsable la Descripción del producto, el equipo de trabajo y el calendario*, pertenecientes al SubProceso de Realización. Por medio de estas tareas, en esta fase donde se ejecuta la técnica y se construye el modelo, a nivel administrativo se distribuye la información a cada responsable, se reúne con el equipo responsable del desarrollo del modelo, se revisa la Descripción del producto, se revisa el cumplimiento del calendario y se realizan correcciones si son necesarios. Como se ha mencionado anteriormente en la explicación del Modelo de Ciclo de Vida propuesto, en esta vuelta del espiral se cicla, donde los distintos prototipos desarrollados y Evaluaciones hechas sobre el modelo juegan un papel importante.

Adicionalmente, en los **Procesos de Desarrollo de Proyecto**, se realiza la actividad de *construir el modelo* propiamente dicho, actividad situada en el SubProceso de Modelado. Mediante esta actividad se setean los parámetros iniciales, se documentan razones de selección, se ejecuta la técnica seleccionada, se procesan los resultados, se describen las características del modelo, se da una explicación detallada del mismo, y se brinda una interpretación del mismo. En esta fase ya se encuentra desarrollado y validado el modelo, por lo que las actividades siguientes tendrán que ver con la Evaluación y Entrega.

La última fase de esta vuelta del espiral es la de **Evaluación del Modelo**. Dentro de los **Procesos de Administración de Proyectos**, específicamente en el SubProceso de Realización, se realizan las siguientes actividades: *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 y revisar los productos terminados durante el proyecto*. Por medio de estas actividades y una vez desarrollado el Modelo, se realizan correcciones si son necesarias, se controla cumplimiento, se realizan ajustes necesarios del plan de adquisición, se revisan subcontratos, se recolectan reportes de actividades y mediciones, se recolectan los productos del trabajo, se revisa el

costo real del proyecto, se revisa el registro de rastreo de los producto, y finalmente se revisan los productos terminados durante el Proyecto y se registran los mismos.

ACTIVIDADES DE LOS PROCESOS	Selección Tecnica Modelado	Generacion Prueba de Diseño	Construccion Modelo	Evaluacion Modelo
Proceso de Administración de Proyectos				
Planificación /Entendimiento del negocio				
- Entendimiento del negocio				
- Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento				
- Definir el protocolo de entrega con el cliente				
- 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 equipo de trabajo				
- Establecer el calendario de actividades				
- Calcular el costo estimado del proyecto				
- Evaluación de la situación				
- Producir un Plan de Proyecto				
- Producir un Plan de Desarrollo				
- Formalizar el inicio de un nuevo ciclo del proyecto				
Realización				
- Acordar las tareas con el equipo de trabajo	■	■	■	
- Acordar la distribución de información		■	■	
- Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario			■	■
- Revisar cumplimiento del plan de adquisiciones y capacitación				■
- 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				
- 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				
- Formalizar la terminación del proyecto o ciclo				
- Llevar a cabo el cierre del contrato con subcontratistas				
- Generar el reporte de mediciones y sugerencias de mejora				
- Planear la entrega				
Proceso de Desarrollo de Proyectos				
Entendimiento del negocio				
- Determinar las metas del Data Mining				
Entendimiento de los datos				
- Reunir los datos iniciales				
- Describir los datos				
- Explorar los datos				
- Verificar la calidad de los datos				
Preparación de los datos				
- Tareas preparatorias				
- Seleccionar los datos				
- Limpiar los datos				
- Construir los datos				
- Integrar los datos				
- Formatear los datos				
Modelado				
- Seleccionar la técnica de modelado	■			
- Generar el diseño de test		■		
- Construir el modelo			■	
- Evaluar el modelo				■
Evaluación				
- Evaluar resultados				
- Revisar el proceso				
- Determinar próximos pasos				
Entrega				
- Producir un reporte final				

Figura 4.8. Mapa de Actividades Propuesto para la quinta vuelta del espiral.

Además, entrando en los **Procesos de Desarrollo de Proyectos**, se realiza la actividad de *evaluar el modelo*, perteneciente al SubProceso de Modelado, por la cual se evalúan los resultados con respecto a los criterios de evaluación, se testean los resultados de acuerdo a la estrategia de testeo, se comparan la evaluación de los resultados y su interpretación, se seleccionan los mejores modelos, se interpretan los resultados en términos de negocios, se chequea el impacto para las metas de E.I., se chequea la confiabilidad de los resultados, se valoran los resultados, y se ajustan los parámetros para obtener un mejor modelo en caso de ser necesario o se cree conveniente.

4.2.2.7. Mapa de Actividades – Última Vuelta del Espiral

Terminando con el análisis del Mapa de Actividades construido anteriormente, se pasa a evaluar la última vuelta del Espiral correspondiente al Modelo de Ciclo de Vida Propuesto. En esta fase se incluyen las siguientes fases: Evaluación Resultados, Revisión Proceso, Determinación Próximos Pasos, Desarrollo Plan, Plan Supervisión y Mantenimiento, Informe Producto y Revisión Proyecto. En la figura 4.9 se puede observar las fases mencionadas y las actividades que se asignan a cada una.

En primer lugar, en la fase de **Evaluación de Resultados** y dentro de los **Procesos de Administración de Proyectos**, se realiza la actividad de *evaluar el cumplimiento del Plan de Proyecto y Plan de Desarrollo*, correspondiente al SubProceso de Evaluación y Control. Por medio de esta actividad se evalúa el cumplimiento del plan de proyecto, se ajusta el mismo, se generan informes, se evalúa el plan de desarrollo y se ajusta el mismo. Estas actividades se desarrollan una vez construido el modelo, por lo que las actividades anteriores no pertenecen a esta fase y fases siguientes tienen más que ver con fases avanzadas de la Evaluación y de la Entrega/Cierre.

Por otro lado, en los **Procesos de Desarrollo de Proyectos**, se lleva a cabo la actividad de *evaluar resultados*, incluida en el SubProceso de Evaluación. En esta actividad se entienden los resultados del Proyecto de E.I., se interpretan los resultados, se evalúan y valoran los resultados, se compara la evaluación, se chequea si existen nuevos objetivos del negocio y finalmente se obtiene la aprobación de los modelos.

En segundo lugar, la fase de **Revisión Proyecto** posee incluida dentro de los **Procesos de Administración de Proyecto**, la actividad de *evaluar el cumplimiento del Plan de Proyecto y Plan de Desarrollo*, correspondiente al SubProceso de Evaluación y Control. En esta actividad se revisa y evalúa el Proyecto y los Planes de los mismos, tanto el de Desarrollo como el de Proyecto en sí.

Teniendo en cuenta los **Procesos de Desarrollo de Proyectos**, se realiza la actividad de *revisar el proyecto* del SubProceso de Evaluación, en la cual se da un examen del proceso de E.I. utilizado, se

analiza el proceso de E.I., se evalúa como puede ser mejorado, se identifican fallas, se identifican pasos faltantes, se identifican alternativas y se revisan resultados del Proyecto de E.I. con respecto a los criterios del éxito del negocio.

La siguiente fase de esta vuelta del espiral es la de **Determinación de los Próximos Pasos**, en la cual por medio de los **Procesos de Administración de Proyectos** y más específicamente en el SubProceso de **Evaluación y Control**, lleva a cabo la tarea de *evaluar el cumplimiento del plan de proyecto y plan de desarrollo*, donde genera informes de ajustes del plan de desarrollo y plan de proyecto, y genera un Reporte de Seguimiento / Plan de Monitoreo y Mantenimiento, previo paso al Cierre y Entrega del Proyecto.

Por otro lado, en los **Procesos de Desarrollo de Proyectos** (SubProceso de **Evaluación**), se incluye la actividad de *determinar próximos pasos*, donde analiza el potencial de despliegue de cada resultado, se estima el potencial de mejora del proceso actual, se chequea recursos sobrantes para determinar si permiten realizar iteraciones del proceso adicionales, se recomiendan alternativas de continuidad, se refina el plan de proceso, se ranquean las posibles acciones y se selecciona y documenta las razones de elección.

La cuarta fase correspondiente a esta última vuelta del Modelo de Ciclo de Vida en Espiral es la de **Desarrollo Plan**. Teniendo en cuenta los **Procesos de Administración de Proyecto**, se define que posee las actividades de *evaluar el cumplimiento del plan de proyecto y plan de desarrollo, analizar y controlar los riesgos, generar el reporte de seguimiento del proyecto* (SubProceso de **Evaluación y Control**) y *formalizar la terminación del proyecto o ciclo, llevar a cabo el cierre del contrato con subcontratistas, generar el reporte de mediciones y sugerencias de mejora, y planear la entrega* (SubProceso **Cierre/Entrega**). Estas actividades generar un Plan de Seguimiento y Plan de Monitoreo y Mantenimiento, documentos de aceptación, informes de subcontratistas, reportes de mediciones y sugerencias de mejoras.

En los **Procesos de Desarrollo de Proyectos**, se lleva a cabo la actividad de *producir un reporte final*, perteneciente al SubProceso de **Entrega**. En esta actividad se identifican que reportes son necesarios para finalizar el Proyecto en cuestión.

En quinto lugar se encuentra la fase de **Plan Supervisión y Mantenimiento**, donde por medio de los **Procesos de Administración de Proyectos**, lleva a cabo las tareas de *recibir y analizar las solicitudes de cambio del cliente, realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos* (SubProceso de **Realización**), *evaluar el cumplimiento del plan de proyecto y plan de desarrollo, analizar y controlar los riesgos, generar el reporte de seguimiento del proyecto* (SubProceso de **Evaluación y Control**) y *formalizar la terminación del proyecto o ciclo, llevar a cabo el cierre del contrato con subcontratistas, generar el reporte de*

mediciones y sugerencias de mejora, y planear la entrega (SubProceso Cierre/Entrega). Por medio de estas tareas se obtiene a nivel administrativo el Plan propiamente dicho que resume la estrategia de supervisión y mantenimiento incluyendo los pasos necesarios y como realizarlos.

Teniendo en cuenta los **Procesos de Desarrollo de Proyectos**, se lleva a cabo la tarea de *producir un reporte final (SubProceso Entrega)*, donde se esboza la estructura del informe, se seleccionan los descubrimientos a ser incluidos en el reporte, se informa la estrategia de supervisión y mantenimiento y se escribe finalmente el reporte.

La siguiente fase del Modelo de Ciclo de Vida es la de **Informe Producto**, donde por medio de los **Procesos de Administración de Proyectos** lleva a cabo las siguientes actividades: *recibir y analizar las solicitudes de cambio del cliente, realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos (SubProceso de Realización)*, *evaluar el cumplimiento del plan de proyecto y plan de desarrollo, analizar y controlar los riesgos, generar el reporte de seguimiento del proyecto (SubProceso de Evaluación y Control)* y *formalizar la terminación del proyecto o ciclo, llevar a cabo el cierre del contrato con subcontratistas, generar el reporte de mediciones y sugerencias de mejora, y planear la entrega (SubProceso Cierre/Entrega)*. Por medio de estas actividades se generan los Reportes de Seguimiento y Mantenimiento, se realiza Informe de SubContratistas, Reportes de Mediciones, se documenta el proceso de E.I. específico, se resumen los detalles para futuros Proyectos similares, se realiza un informe de los contratos, se comunica la finalización a todos los interesados, entre otras tareas.

Teniendo en cuenta los **Procesos de Desarrollo de Proyectos**, se lleva a cabo la actividad de *producir un reporte final* perteneciente al SubProceso de Entrega. En esta actividad se escribe el reporte e informe final y se lleva a cabo la presentación final del Proyecto.

La última fase del Modelo de Ciclo de Vida en Espiral propuesto en este capítulo y que se analiza con respecto al Mapa de Actividades construido es la de **Revisión Proyecto**. Teniendo en cuenta los **Procesos de Administración de Proyectos**, se llevan a cabo las siguientes tareas: *recibir y analizar las solicitudes de cambio del cliente, realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos (SubProceso de Realización)*, *evaluar el cumplimiento del plan de proyecto y plan de desarrollo, analizar y controlar los riesgos, generar el reporte de seguimiento del proyecto (SubProceso de Evaluación y Control)* y *formalizar la terminación del proyecto o ciclo, llevar a cabo el cierre del contrato con subcontratistas, generar el reporte de mediciones y sugerencias de mejora, y planear la entrega (SubProceso Cierre/Entrega)*. Por medio de estas actividades se generan los reportes e informes finales y se lleva a cabo el Plan de Entrega, donde se planifica la entrega de los productos generados.

ACTIVIDADES DE LOS PROCESOS	Evaluación Resultados	Revisión Proceso	Determinación Proximo Pasos	Desarrollo Plan	Plan Supervision y Mantenimiento	Informe Producto	Revisión Proyecto
Proceso de Administración de Proyectos							
Planificación /Entendimiento del negocio							
- Entendimiento del negocio							
- Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento							
- Definir el protocolo de entrega con el cliente							
- 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 equipo de trabajo							
- Establecer el calendario de actividades							
- Calcular el costo estimado del proyecto							
- Evaluación de la situación							
- Producir un Plan de Proyecto							
- Producir un Plan de Desarrollo							
- Formalizar el inicio de un nuevo ciclo del proyecto							
Realización							
- Acordar las tareas con el equipo de trabajo							
- Acordar la distribución de información							
- Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario							
- Revisar cumplimiento del plan de adquisiciones y capacitación							
- 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							
- 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							
- Formalizar la terminación del proyecto o ciclo				■	■	■	■
- Llevar a cabo el cierre del contrato con subcontratistas				■	■	■	■
- Generar el reporte de mediciones y sugerencias de mejora				■	■	■	■
- Planear la entrega				■	■	■	■
Proceso de Desarrollo de Proyectos							
Entendimiento del negocio							
- Determinar las metas del Data Mining							
Entendimiento de los datos							
- Reunir los datos iniciales							
- Describir los datos							
- Explorar los datos							
- Verificar la calidad de los datos							
Preparación de los datos							
- Tareas preparatorias							
- Seleccionar los datos							
- Limpiar los datos							
- Construir los datos							
- Integrar los datos							
- Formatear los datos							
Modelado							
- Seleccionar la técnica de modelado							
- Generar el diseño de test							
- Construir el modelo							
- Evaluar el modelo							
Evaluación							
- Evaluar resultados	■						
- Revisar el proceso		■					
- Determinar próximos pasos			■				
Entrega							
- Producir un reporte final				■	■	■	■

Figura 4.9. Mapa de Actividades Propuesto para la ultima vuelta del espiral.

Con respecto a los **Procesos de Desarrollo de Proyectos**, se realiza la actividad de *producir un reporte final* perteneciente al SubProceso de Entrega, donde se lleva a cabo la Presentación Final del Proyecto y donde se documenta todo el desarrollo del Proyecto para futuras experiencias y trabajos similares a realizar.

En este capítulo se ha presentado la solución propuesta para los problemas mencionados en el capítulo anterior. En primer lugar se llevo a cabo la propuesta de un Modelo de Ciclo de Vida en forma de Espiral, describiendo cada vuelta y cada fase del Modelo presentado e informando el valor agregado que este Modelo le ha dado a los Modelos de Ciclos de Vida existentes para Proyectos de Explotación de Información. Posteriormente se llevo a cabo la construcción de un Mapa de Actividades teniendo como ejes en primer lugar el Modelo de Procesos para Proyectos de este tipo desarrollado por Vanrell [2012] y en segundo lugar las fases del Modelo de Ciclo de Vida propuesto en este trabajo.

En el siguiente capítulo se lleva a cabo la generación de una prueba de concepto donde se pone en práctica lo desarrollado en este capítulo.

5. Pruebas de Concepto

En este capítulo se lleva a cabo la prueba de las soluciones propuestas en el capítulo anterior. Para ello se presenta un caso de prueba donde se aplica la solución propuesta. En primer lugar, en la sección 5.1., se presenta el Proyecto de Explotación de Información seleccionado para realizar dicha prueba de concepto. En segundo lugar, se lleva a cabo la elección de un Modelo de Ciclo de Vida que sea adecuado a dicho Proyecto (sección 5.2.). Por último, en la sección 5.3., se pone en marcha el desarrollo del Proyecto de Explotación de Información teniendo en cuenta el Modelo de Ciclo de Vida seleccionado y las actividades definidas en el Mapa de Actividades construido en el capítulo anterior.

5.1. Presentación de Proyecto para Prueba de Concepto

En esta sección se presenta el Proyecto de Explotación de Información que se toma para realizar la Prueba de Concepto de lo propuesto en el capítulo anterior, el cual tiene como fin encontrar patrones de comportamientos para detectar enfermedades. Para ello se poseen datos de pacientes con síntomas de alguna enfermedad hepática de ciudades como Salta, Jujuy, Formosa y Chaco. A un porcentaje se le ha detectado que posee alguna enfermedad y a otros no.

La detección temprana de alguna enfermedad es beneficiosa para el paciente, ya que se puede atacar al problema de manera más rápida y no cuando ya el mismo se encuentra avanzado.

Como base para realizar este Proyecto de Explotación de Información se tomo un archivo con 583 casos de pacientes que se han tratado en el Norte de la República Argentina (a 416 se le detectaron enfermedad y al resto no). Para dicho diagnostico fueron necesarios varios atributos y variables, cuyos valores determinan el diagnostico final que se le da al paciente. El Proyecto utiliza estos valores para poder armar reglas que le sirvan a cualquier organismo de Salud.

5.2. Elección de Modelo de Ciclo de Vida para el Proyecto

Para llevar a cabo la prueba de concepto y el desarrollo de este Proyecto de E.I. descrito brevemente anteriormente resulta necesario en primer lugar realizar la elección de un Modelo de Ciclo de Vida. Como alternativas para esta elección se encuentran 2 modelos: la metodología CRISP-DM y el Modelo de Ciclo de Vida propuesto en este Trabajo Final de Licenciatura en el capítulo anterior.

El Modelo de Ciclo de Vida propuesto ha tomado como base el Modelo de Procesos desarrollado por Vanrell [2012], el cual aporta una mejora a la metodología CRISP-DM agregándole los Procesos correspondientes a la Administración del Proyecto, los cuales resultan necesarios para

recolectar información que le sea útil a los procesos de Desarrollo para aumentar la calidad de los mismos permitiendo realizar ajustes y mantener un estándar en la realización de distintos proyectos llevando a cabo la conversión de procesos caóticos y aislados en procesos repetitivos y controlados, tal como se describió en el capítulo 2. Además de esta mejora y este adicional que posee el Modelo de Procesos seleccionado y por lo cual, también posee el Modelo de Ciclo de Vida propuesto en este trabajo, éste último también le agrega a la metodología CRISP-DM un análisis enfocado a los riesgos. Este análisis y tratamiento de los riesgos no es custodiado por CRISP_DM y sí por este Modelo de Ciclo de Vida Propuesto.

Por todas estas razones, se cree conveniente para llevar a cabo el desarrollo de este Proyecto de Explotación de Información, y que sirve como Prueba de Concepto de lo propuesto en este Trabajo, la utilización del Modelo de Ciclo de Vida presentado en este Trabajo Final de Licenciatura y el Mapa de Actividades construido en el capítulo anterior con eje al Modelo de Procesos desarrollado por Vanrell [2012].

5.3. Desarrollo del Proyecto de Explotación de Información

En esta sección se lleva a cabo el desarrollo de un Proyecto de Explotación de Información con base al Modelo de Ciclo de Vida propuesto y al Mapa de Actividades construido con eje al Modelo de Ciclo de Vida y al Modelo de Procesos de Vanrell. Las actividades que posee cada fase fueron descritas en el capítulo anterior, donde se justificó su elección y donde se menciona que áreas no se realizan y porqué.

A continuación se lleva a cabo el desarrollo del Proyecto de Explotación de Información teniendo en cuenta el Mapa de Actividades propuesto.

5.3.1. Procesos de Administración de Proyectos

En esta sección se llevan a cabo el desarrollo de cada SubProceso correspondiente a la Administración de Proyectos, explicando por la fase del Modelo de Ciclo de Vida en Espiral que cada actividad pasa.

5.3.1.1. Entendimiento del Negocio

Este subproceso está compuesto por las siguientes secciones: Determinar los objetivos del negocio, Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento, Definir el protocolo de entrega con el cliente, 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 equipo de trabajo, Establecer

el calendario de actividades, Calcular el costo estimado del proyecto, Evaluación de la situación, Producir un Plan de Proyecto, Producir un Plan de Desarrollo, y Formalizar el inicio de un nuevo ciclo del proyecto.

5.3.1.1.1. Determinar los Objetivos del Negocio

Esta actividad se lleva a cabo en la fase de Objetivos del Negocio, Evaluación de la Situación y Objetivos del Proyecto de E.I., pertenecientes a la primera y segunda vuelta del Modelo de Ciclo de Vida en Espiral propuesto. Como se ha mencionado en el capítulo anterior, estas actividades se llevan a cabo en estas fases dado que la salida correspondiente a estas tareas son la Base de Conocimiento del Negocio; los Objetivos del Negocio, describiendo el problema en términos generales, evaluando el estado actual del proyecto, especificando los requerimientos y preguntas del negocio, determinando los beneficios esperados en términos de negocio; y los Criterios del Éxito del negocio. A continuación se lleva a cabo estas actividades teniendo en cuenta el Proyecto de Explotación de Información seleccionado para realizar la Prueba de Concepto del Trabajo Final de Licenciatura:

5.3.1.1.1.1. Base de conocimiento

La consultora Consulting S.A. es una organización privada dedicada a desarrollos de sistemas tanto convencionales como sistemas de explotación de información. A través de distintos clientes, esta consultora produce una gran gama de sistemas de explotación de información los cuales dotan a las compañías de conocimiento necesario para poder tomar decisiones más acertadas. En este trabajo, el cliente que solicita un proyecto de este tipo es el Ministerio de Salud de la República Argentina, para determinar mediante una serie de datos y diagnósticos, que características son más significativas para determinar enfermedades hepáticas y con cuales características no se está enfermo.

La consultora cuenta actualmente con un total de 85 empleados, entre directivos y personal de otros sectores más allá del departamento de Informática. La empresa cuenta con una dirección ejecutiva dirigida por el Dr. Eduardo Díaz. La organización se encuentra dividida en 4 departamentos: Departamento de Control de Gestión (Manuel Alonso como responsable), Departamento de Informática (Joel López como responsable), Departamento de Auditoria (a cargo de Alfredo López) y Departamento Administrativo (dirigido por Guillermo Rossi). El Departamento de Control de Gestión posee el área contable, a cargo de Ana Pio. En cambio, el Departamento de Informática está dividido en el Área de Sistemas (Lisandro Fredes), Área de Desarrollo (Matías Méndez) y de Investigación (Ezequiel Más). Por otro lado el Departamento Administrativo posee el Área Legal y el Área Administrativa. A continuación se presenta el Organigrama Final de la Organización:



Fig. 5.1. Organigrama de la Organización

Por otro lado, el Ministerio de Salud de la República Argentina (el cliente en cuestión), esta dirigido por el Dr. Manzur y posee 4 Secretarías: de Políticas, Regulación e Institutos (Yedlin), de Promoción y Programas Sanitarios (Diosque), de Determinantes de la Salud y Relaciones Sanitarias (Villa) y la Secretaría de Coordinación (Garvich). Como subsecretaria se tienen: Subsecretaria de Políticas, Regulación y Fiscalización, Subsecretaria de Gestión de Servicios Asistenciales (Secretaría de Políticas, Relación e Institutos); Subsecretaria de Salud Comunitaria y Subsecretaria de Prevención y Control de Riesgos (Secretaría de Promoción y Programas Sanitarios); Subsecretaria de Relaciones Sanitarias e Investigación (Secretaría de Determinantes de la Salud y Relaciones Sanitarias); y Subsecretaria de Coordinación Administrativa (Secretaría de Coordinación).

La Subsecretaria que contrata los servicios de la Consultora para llevar a cabo el Proyecto de Explotación de Información sobre Enfermedades del Hígado es la Subsecretaria de Relaciones Sanitarias e Investigación.

Para poder llevar a cabo este Proyecto de Explotación de Información (EI), resultan personas claves el director de la Consultora Consulting Eduardo Díaz, el Ing. Joel López como responsable del Departamento de Informática y los responsables de las Áreas de Sistemas, Desarrollo e Investigación (Lisandro Fredes, Matías Méndez y Ezequiel Mas). Como contacto del Ministerio de Salud se tiene al Dr. Lazovski, el cual solicita los servicios de la Consultora y brinda los requerimientos y problema en cuestión.

El Departamento de Gestión de la Consultora es quien lleva a cabo la contaduría de la organización. Mediante el monto presupuestado al Ministerio de Salud se poseen los recursos financieros necesarios para poder encarar exitosamente este proyecto. El presupuesto es consensuado entre el Director del Departamento de Gestión y el Director de Informática, para ser finalmente aprobado por el Director General.

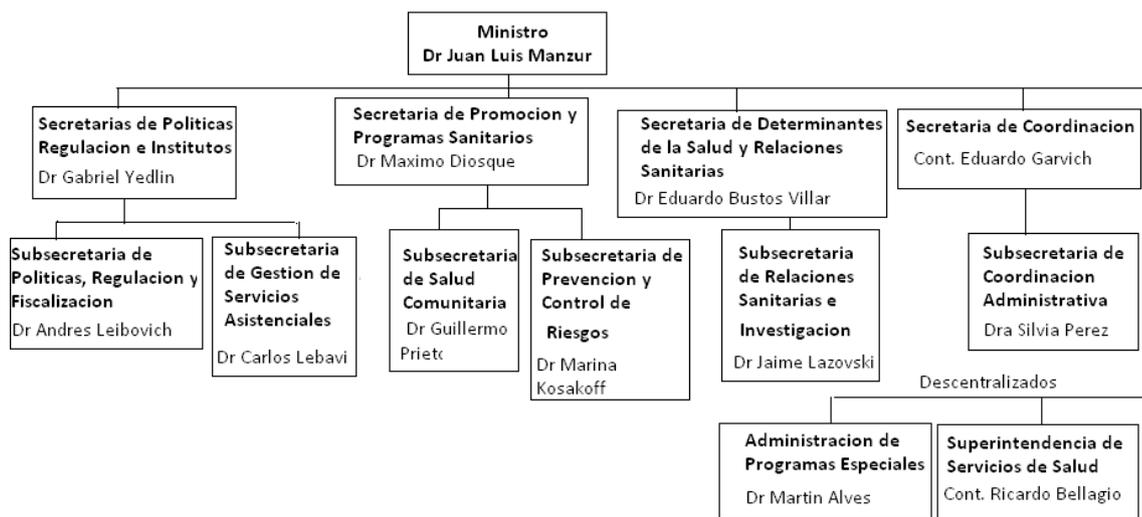


Fig. 5.2. Organigrama del Ministerio de Salud

El Gobierno Nacional de la República Argentina es quien, a través del Ministerio de Salud, quien solicita el desarrollo de este Proyecto de EI, para poder abastecerlos de conocimientos necesarios en cuanto a los pacientes del hígado.

Los departamentos involucrados en este Negocio dentro de la Consultora son el Departamento de Gestión (llevan a cabo la gestión de la organización, la cual es necesaria contarla en este proyecto), el Departamento Administrativo (administra los Recursos de la Organización) y, por sobre todo, el Departamento de Informática (quienes finalmente van a desarrollar este proyecto).

5.3.1.1.1.2. Objetivos del Negocio

El hígado es el órgano más grande dentro del cuerpo. También es uno de los más importantes. El hígado tiene muchas funciones, incluyendo la transformación de los alimentos en energía y la eliminación del alcohol y las toxinas de la sangre. El hígado también produce bilis, un líquido amarillo verdoso que ayuda a la digestión.

Existen muchos tipos de enfermedades hepáticas. Algunas de ellas son causadas por virus, como la hepatitis A, la hepatitis B y la hepatitis C. Otras pueden ser a consecuencia de medicamentos, venenos o toxinas o por ingerir demasiado alcohol. Si el hígado forma tejido cicatricial por una enfermedad, se denomina cirrosis. La ictericia, o coloración amarilla de la piel, puede ser un signo de enfermedad hepática.

Al igual que en otras partes del cuerpo, el cáncer de hígado puede afectar el hígado. Otras enfermedades hepáticas pueden ser hereditarias, como por ejemplo, la hemocromatosis.

Dada la gran cantidad de pacientes en el último tiempo con síntomas de enfermedades hepáticas, se necesita un sistema de Explotación de Información, con datos tomados del norte del país, en

provincias como Salta, Jujuy, Formosa y Chaco, sobre pacientes con síntomas similares, para determinar con más precisión y certeza cuales son los que realmente están enfermos.

Los resultados de estos estudios son muy importantes para el desarrollo del sistema automático de diagnóstico médico.

Actualmente es de gran preocupación para el Ministerio de Salud de la República Argentina, el incremento de pacientes que poseen síntomas de enfermedades hepáticas. Por ello, se tomo como base los datos provenientes del norte del país con más de 550 casos de estudios y los factores que determinan que el paciente finalmente posea o no alguna enfermedad de este estilo.

Gracias a este Proyecto, las distintas entidades de Salud a lo largo y ancho del país van a poder diagnosticar con mayor precisión.

Para poder llevar a cabo este proyecto es necesario contar con una serie de datos que sirva para analizar y sacar conclusiones, grandes conocimientos del tema a abordar, conocimientos suficientes por parte de los que van a desarrollar este proyecto en Tecnologías de Explotación de Información, recursos necesarios para poder realizar este plan (financieros, humanos, materiales, etc.) y conocimiento de las Herramientas actuales para poder elegir la mejor para poder llevar a cabo un Proyecto de calidad, en los plazos definidos.

La Explotación de Información es la sub-disciplina Informática que aporta a la Inteligencia de Negocio las herramientas para la transformación de información en conocimiento. Se ha definido como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información. Al hablar de explotación de información basada en sistemas inteligentes se refiere específicamente a la aplicación de métodos de sistemas inteligentes, para descubrir y enumerar patrones presentes en la información. Los métodos basados en sistemas inteligentes, permiten obtener resultados de análisis de la masa de información que los métodos convencionales no logran tales como: los algoritmos TDIDT (Top Down Inducción Decisión Trees), los mapas auto organizados (SOM) y las redes bayesianas. Los algoritmos TDIDT permiten el desarrollo de descripciones simbólicas de los datos para diferenciar entre distintas clases. Los mapas auto organizados pueden ser aplicados a la construcción de particiones de grandes masas de información. Tienen la ventaja de ser tolerantes al ruido y la capacidad de extender la generalización al momento de necesitar manipular datos nuevos. Las redes bayesianas pueden ser aplicadas para identificar atributos discriminantes en grandes masas de información, detectar patrones de comportamiento en análisis de series temporales.

Los destinatarios finales de este Proyecto, o quienes se van a nutrir de lo desarrollado por la Consultora Consulting es el Ministerio de Salud de la República Argentina. Finalmente, los que se van a nutrir de conocimiento obtenido por este sistema son las distintas entidades de salud del país,

los cuales van a poder diagnosticar con mayor precisión cuando un paciente posee una enfermedad hepática y cuando los síntomas no son suficientes para detectar este diagnóstico.

Las expectativas de los usuarios finales es poder abastecerse de conocimiento necesario para poder tomar decisiones y diagnosticar con mayor precisión. Se han de formular las siguientes preguntas del Negocio:

- Que son las enfermedades hepáticas?
- Cuales son ellas?
- Cuáles son los síntomas más significativas para diagnosticar alguna enfermedad del hígado?

Como principales beneficios esperados de este proyecto se obtienen nutrir a los hospitales, clínicas, médicos, doctores y distintos organismos de salud de conocimiento para que puedan detectar de manera más rápida y eficiente cuando un paciente posee alguna enfermedad hepática y cuando no. La detección temprana de alguna enfermedad es beneficiosa para el paciente, ya que se puede atacar al problema de manera más rápida y no cuando ya el mismo se encuentra avanzado.

5.3.1.1.1.3. Criterios del Éxito

El éxito de este proyecto estará alcanzado si el Ministerio de Salud puede nutrir a sus organismos de información y conocimiento necesario para que puedan tomar decisiones de manera precisa y que muchos pacientes puedan ser diagnosticados de manera exacta y sin errores que pueden acarrear más problemas.

Los criterios del éxito los determina principalmente el cliente de la consultora, o sea el Ministerio de Salud de la República Argentina, el cual solicita dicho proyecto y financia el mismo.

5.3.1.1.2. Definir el proceso específico basado en la descripción del proyecto y el proceso de desarrollo y mantenimiento

Esta actividad perteneciente al SubProceso de Planificación /Entendimiento del Negocio, se lleva a cabo en la Fase Objetivos del Proyecto de Explotación de Información correspondiente a la segunda vuelta del Modelo de Ciclo de Vida en Espiral propuesto, dado que, como se explico en el capítulo anterior, tareas tales como evaluar la descripción del proyecto, evaluar el proceso de desarrollo y mantenimiento, describir el problema en términos generales, clasificar prerequisites del proyecto, especificar los beneficios esperados y definir el proceso específico a utilizar, son incluidas dentro de esta fase. A continuación se realiza la actividad mencionada para el Proyecto en cuestión:

El departamento de Informática y los desarrolladores de este proyecto de EI poseen un conjunto de datos tomados de pacientes del Norte de Argentina, los cuales son útiles y suficientes para poder desarrollar reglas de comportamiento de síntomas y cuales pueden provocar en mayor o menor

medida la aparición de alguna enfermedad hepática. El Ministerio de Salud de la República Argentina ha brindado los requisitos que este proyecto posee.

La consultora Consulting documento los requisitos brindados por el Ministerio de Salud y analizo los pasos a seguir en cuanto al desarrollo. Por medio de herramientas y algoritmos de explotación de información conocidos por los desarrolladores, se realizará dicho proyecto. Las fases de desarrollo se describirán en la sección que se describirá a continuación

5.3.1.1.3. Definir ciclos y actividades con base en la descripción del proyecto y en el proceso específico

Esta actividad se realiza en las fases de Plan de Proyecto y Validación de Objetivos correspondientes a la segunda vuelta del Modelo de Ciclo de Vida en Espiral propuesto en el capítulo anterior y a continuación se lleva a cabo para el Proyecto de E.I. en cuestión:

A continuación se mencionara cada ciclo del proyecto:

- Análisis requisitos
- Obtención de los Datos
- Análisis de Datos
- Limpieza de Datos
- Determinación de técnicas y algoritmos a utilizar
- Ejecución
- Análisis de resultados

Las actividades principales para cada ciclo mencionado en la sección anterior son:

- Análisis Requisitos: Analizar los requerimientos y determinar objetivos
- Obtención los Datos: Conseguir una colección de datos reales los cuales sirvan para poder producir reglas. Estos mismos deben ser significativos y debe haber una gran variedad
- Análisis los Datos: Llevar a cabo un análisis profundo de los Datos a procesar para que los mismos sean útiles al llevar a cabo la ejecución del programa.
- Limpieza los Datos: Llevar a cabo una limpieza o depuración de datos que no colaboren con el desarrollo del proyecto
- Determinación de técnicas y algoritmos a utilizar: Analizar las herramientas y algoritmos existentes en torno a procesos de explotación de información para seleccionar los más adecuados para este proyecto y conseguir los objetivos mencionados.
- Ejecución: Ejecutar el proceso y el algoritmo sobre los datos seleccionado.
- Análisis de resultados: Analizar los resultados obtenidos y verificar que los requerimientos del usuario son satisfechos.

5.3.1.1.4. Determinar tiempo estimado para cada actividad

La presente actividad, tal como se menciona en el capítulo anterior, se lleva a cabo en las fases de Plan de Proyecto y Validación de Objetivos y a continuación se realiza para el presente Proyecto de Explotación de Información:

La estimación de realización de cada ciclo y actividad de desarrollo es realizada por el Project Manager de la Consultora, teniendo la aprobación del Director de la Consultora y del encargado del Departamento de Informática.

A continuación se presenta el GANTT para la estimación de tiempo de cada tarea:

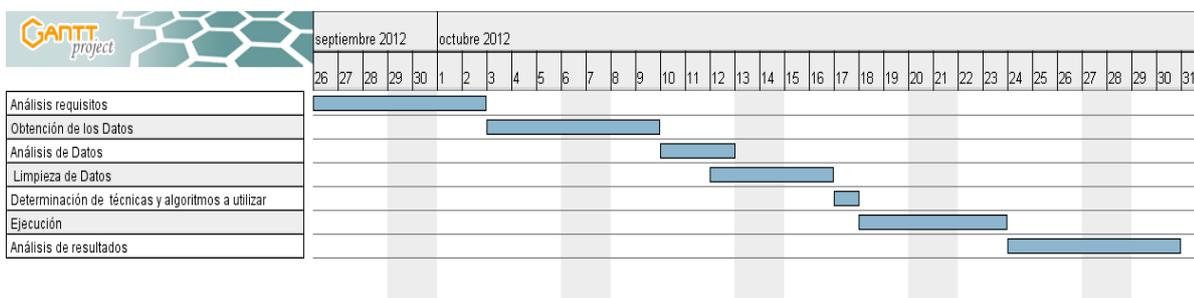


Fig. 5.3. GANTT de los ciclos

5.3.1.1.5. Elaborar plan de adquisiciones y capacitación

Al igual que la actividad anterior perteneciente al SubProceso de Planificación / Entendimiento del Negocio esta actividad se lleva a cabo en la Fase de Plan de Proyecto y en la de Validación de Objetivos.

Para la realización de dicho proyecto es necesario contar con un encargado que determine los objetivos y requerimientos y una persona con conocimientos en tecnologías de explotación de información a fin de desarrollar el proyecto en cuestión. Como recursos se necesitan herramientas de explotación de información (Tanagra), una computadora, datos y un enunciado para poder entender a que se quiere llegar y que se quiere lograr.

Previo al desarrollo en si del proyecto es necesaria una capacitación en cuanto al problema en cuestión, la necesidad de contar con este conocimiento a generar, los problemas actuales de salud y enfermedades hepáticas actuales. Posteriormente, una vez entendido el problema y la necesidad actual por la que se solicito este proyecto, es necesaria una capacitación de limpieza de datos, para poder determinar de forma correcta que datos se van a suprimir y porque. Para cada ciclo es necesario contar con algún recurso, ya se humano como material.

5.3.1.1.6. Establecer el equipo de trabajo

Esta actividad se realiza de igual manera en las fases de Plan de Proyecto y Validación de Objetivos pertenecientes a la segunda vuelta del Modelo de Ciclo de Vida en Espiral. A continuación se lleva a cabo para el Proyecto en cuestión:

El encargado del Departamento de Informática y el Project Manager son quienes tienen una entrevista con el cliente y documentan los requisitos. El director de la Consultora aprueba el desarrollo del proyecto y el mismo se pone en ejecución. Dentro del Departamento de Informática, el Project Manager es quien se encarga de que el desarrollo se haga en el tiempo y la forma estipulada. Dentro del equipo de Investigación, hay un investigador el cual analiza los problemas y obtiene los datos. Dentro del equipo de desarrollo, hay un desarrollador el cual se encarga de programar el proyecto de explotación de información, informándole los resultados al Project Manager.

5.3.1.1.7. Establecer el calendario de actividades

Esta actividad cubre a las fases de Plan de Proyecto y a la de Validación de Objetivos.

Las fechas estipuladas de realización de cada actividad fueron descritas en el Diagrama GANNT de la sección 5.3.1.1.4.

5.3.1.1.8. Calcular el costo estimado del proyecto

El cálculo de los costos mantiene a las fases de Plan de Proyecto y Validación de Objetivos de la segunda vuelta del Modelo de Ciclo de Vida.

Los costos son provenientes a los salarios de cada involucrado dentro del proyecto, sumado al consumo de luz y otros servicios utilizados durante el desarrollo del mismo. Dado que los datos fueron provistos por una página Web y que la herramienta Tanagra a utilizar es gratuita, no se estiman costos en estos aspectos. El valor total del proyecto es de aproximadamente \$50.000.- y es financiado totalmente por el Ministerio de Salud.

5.3.1.1.9. Evaluación de la situación

La evaluación de la situación se lleva a cabo en las fases de Evaluación de Situación propiamente dicho, Plan de Proyecto y Validación de Objetivos pertenecientes a la segunda vuelta del Modelo de Ciclo de Vida en Espiral, y a todas las fases tanto de la tercera como de la cuarta vuelta del Espiral. A continuación se evalúa la situación para el Proyecto de E.I. actual:

Como recurso hardware disponible se posee una Notebook Samsung 900X3A, con 4GB de memoria y un procesador Intel® Core™ i5 Processor 2537M (1,40 GHz, 3 MB). Por medio de esta computadora el desarrollador llevara a cabo este proyecto.

Como herramienta técnica para llevar a cabo este proyecto de explotación de información se posee la herramienta Tanagra (<http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>). Además se utilizara el modelo de Procesos propuesto por Vanrell, el cual tiene como base la metodología CRISP-DM.

Las restricciones encontradas para la realización de este proyecto son los plazos de tiempos. El acceso a los datos y descripción de cada campo está permitido y no se infringe ninguna ley. El conocimiento relevante es accesible ya que la descripción del problema esta transparente y los conocimientos sobre proyectos de explotación de información fueron asentados en este periodo. Por otro lado no se encuentran restricciones en cuanto al presupuesto

El riesgo del negocio encontrado es la aparición de cambios en los requerimientos por parte del cliente. En cambio, el riesgo organizacional que podría llegar a presentarse es la ida de algún miembro del equipo involucrado en este proyecto

No se encuentran riesgos financieros ya que la herramienta a utilizar es gratuita y el presupuesto del proyecto abarca los salarios de los involucrados.

Los riesgos técnicos encontrados son que los recursos materiales de la organización sufran algún siniestro.

Otro riesgo posible es que los datos no sean lo suficientemente buenos como para poder realizar este proyecto

Riesgo	Acción
Aparición de cambios en los requerimientos por parte del cliente.	Entrevista más profunda para no sufrir cambios de nuevo. Cambios en los tiempos estipulados si el requerimiento fue cambiado por cliente.
El riesgo organizacional que podrían llegar a presentarse es la ida de algún miembro del equipo involucrado en este proyecto	Capacitación a más de un recurso humano para que el cambio no sea atroz
Los riesgos técnicos encontrados son que los recursos materiales de la organización sufran algún siniestro.	Almacenamiento de recursos materiales adicionales para suplir a los dañados.
Datos no sean lo suficientemente buenos como para poder realizar este proyecto.	Búsqueda de varias fuentes de datos para obtener alternativas y gran variedad.

Table 5.1. Plan de Contingencia

Al analizar los datos obtenidos surgieron algunos términos los cuales es necesario entender para poder lleva a cabo este proyecto. Los siguientes términos son útiles para comprender el problema en cuestión:

- **Bilirrubina:** pigmento biliar de color amarillo anaranjado que resulta de la degradación de la hemoglobina.
- **Hemoglobina:** heteroproteína de la sangre, de masa molecular 64.000 g/mol (64 kDa), de color rojo característico, que transporta el oxígeno desde los órganos respiratorios hasta los tejidos, el dióxido de carbono desde los tejidos hasta los pulmones que lo eliminan y también participa en la regulación de pH de la sangre, en vertebrados y algunos invertebrados
- **Fosfatasa alcalina:** enzima hidrolasa responsable de eliminar grupos de fosfatos de varios tipos de moléculas como nucleótidos, proteínas y alcaloides.
- **Alanina aminotransferasa:** enzima aminotransferasa con gran concentración en el hígado y en menor medida en los riñones, corazón y músculos.
- **Enzima:** moléculas de naturaleza proteica que catalizan reacciones químicas
- **Aspartato aminotransferasa:** enzima aminotransferasa que se encuentra en varios tejidos del organismo de los mamíferos, especialmente en el corazón, el hígado y el tejido muscular.
- **Proteínas:** macromoléculas formadas por cadenas lineales de aminoácidos.
- **Albumina:** proteína que se encuentra en gran proporción en el plasma sanguíneo, siendo la principal proteína de la sangre y una de las más abundantes en el ser humano. Es sintetizada en el hígado
- **Globulina:** proteínas solubles en agua que se encuentran en todos los animales y vegetales

Los criterios de selección de técnicas y herramientas son que cubran las necesidades del proyecto, permitan gran cantidad de dato, resultados fáciles de visualizar, fácil uso, bajo costo y buena performance. Dado el conocimiento y experiencia que se posee en la herramienta Tanagra, se eligió esta herramienta para llevar a cabo el desarrollo del Proyecto. Tanagra es ágil y fácil de usar, con un potencial muy alto y posee la ventaja que su utilización es gratuita, lo cual colabora con el financiamiento del proyecto.

5.3.1.1.10. Producir un Plan de Proyecto

Esta actividad se lleva a cabo y mantiene a las fases de Plan de Proyecto y Validación de Objetivos de la segunda vuelta del Modelo de Ciclo de Vida en Espiral propuesto y, además, en la Validación de Datos correspondiente a la cuarta vuelta del espiral. A continuación se lleva a cabo el Plan de Proyecto para el Proyecto actual usado como prueba de concepto:

La consultora ha identificado 7 ciclos sobre los cuales se va a llevar a cabo este proyecto de Explotación de Información: Análisis Requisitos, Obtención los Datos, Análisis los Datos, Limpieza los Datos, Determinación de técnicas y algoritmos a utilizar, Ejecución, Análisis de

resultados. Para cada tarea se han identificado una serie de actividades principales a realizar en cada ciclo:

- **Análisis Requisitos:** Analizar los requerimientos y determinar objetivos
- **Obtención los Datos:** Conseguir una colección de datos reales los cuales sirvan para poder producir reglas. Estos mismos deben ser significativos y debe haber una gran variedad
- **Análisis los Datos:** Llevar a cabo un análisis profundo de los Datos a procesar.
- **Limpieza los Datos:** Llevar a cabo una limpieza o depuración de datos que no colaboren con el desarrollo del proyecto
- **Determinación de técnicas y algoritmos a utilizar:** Analizar las herramientas y algoritmos existentes en torno a procesos de explotación de información para seleccionar los más adecuados para este proyecto y conseguir los objetivos mencionados.
- **Ejecución:** Ejecutar el proceso y el algoritmo sobre los datos
- **Análisis de resultados:** Analizar los resultados obtenidos y verificar que los requerimientos del usuario son satisfechos.

En cuanto a estimaciones, se ha mencionado que no se encontraron costos en la obtención de los datos. Por otro lado, hay que tener en cuenta los salarios de cada involucrado dentro del proyecto, sumado al consumo de luz y otros servicios utilizados durante el desarrollo del mismo. Dados que los datos fueron provistos por una página Web y que la herramienta Tanagra a utilizar es gratuita, no se estiman costos en estos aspectos. Por otro lado, se ha estimado en cuanto al tiempo de realización de dicho proyecto. Se ha pactado como fecha de entrega fines de Octubre, por lo que se dividió cada ciclo en periodos de tiempo para ir avanzando en los tiempos esperados. A continuación se presenta el siguiente diagrama GANTT con los días esperados de desarrollo de cada ciclo y su esperado fin:

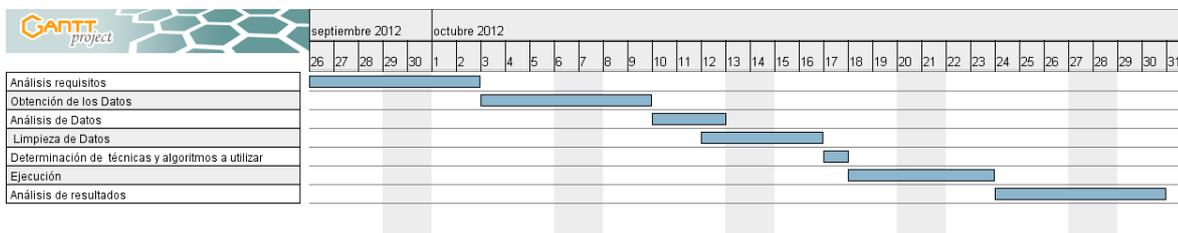


Fig. 5.4. GANTT de los ciclos

Como puede observarse se espera que el proyecto finalice el 31/10, teniendo aproximadamente más de 1 mes para su realización.

El equipo de trabajo está armado por el Lic. Matias Mendez en el Área de Desarrollo, el Lic Mas en el Área de Investigación, el Project Manager (dentro del área de desarrollo) Mauro Matos, el

director del departamento de Informática Joel Lopez y el Director de Consulting Eduardo Díaz. El programador dentro del Área de Desarrollo es el Sr. Omar Pérez.

Por otro lado se han presentado una serie de riesgos que pueden llegar a ocurrir en el proyecto y al mismo tiempo se ha presentado un plan de contingencias para cada riesgo. A continuación se presenta un cuadro donde se muestra el Riesgo posible y la acción a realizar:

La entrega se realizara en el tiempo estipulado, se hará en las oficinas del Ministerio de Salud del País y se le invitara al usuario final que pruebe el proyecto y evalué lo realizado.

Riesgo	Acción
Aparición de cambios en los requerimientos por parte del cliente.	Entrevista más profunda para no sufrir cambios de nuevo. Cambios en los tiempos estipulados si el requerimiento fue cambiado por cliente.
El riesgo organizacional que podrían llegar a presentarse es la ida de algún miembro del equipo involucrado en este proyecto	Capacitación a más de un recurso humano para que el cambio no sea atroz
Los riesgos técnicos encontrados son que los recursos materiales de la organización sufran algún siniestro.	Almacenamiento de recursos materiales adicionales para suplir a los dañados
Datos no sean lo suficientemente buenos como para poder realizar este proyecto	Búsqueda de varias fuentes de datos para obtener alternativas y gran variedad.

Table 5.2. Plan de Contingencia

5.3.1.1.11. Formalizar el inicio de un nuevo ciclo del proyecto

La última actividad del SubProceso Planeamiento / Entendimiento del Negocio correspondiente a los Procesos de Administración de Proyecto, es útil y se lleva a cabo en las fases de Plan de Proyecto y Validación de Objetivos (segunda vuelta del espiral) y en la Validación de Datos. Siguiendo con el desarrollo del Proyecto de E.I., se realiza la actividad mencionada:

El Proyecto se ha iniciado formalmente el día Miércoles 26/10/2012 y se le ha notificado a cada integrante e involucrado directa o indirectamente en el proyecto del inicio del mismo.

5.3.1.2. Realización

Este SubProceso posee las siguientes actividades que se llevaran a cabo para el actual Proyecto de Explotación de Información: 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, Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos

5.3.1.2.1. Acordar las tareas con el equipo de trabajo

Esta primera actividad de este SubProceso, correspondiente a los Procesos de Administración de Proyectos, se lleva a cabo en las fases Plan de Proyecto y Validación de Objetivos (segunda vuelta del espiral), y Selección de Técnica de Modelado, Generación Prueba de Diseño y Construcción del Modelo (quinta vuelta del espiral). A continuación se realiza para el Proyecto actual:

Se han revisado las tareas a realizar en el ciclo actual. Según la estimación de tiempo y lo descrito en el Diagrama GANTT de las secciones anteriores, para este periodo de tiempo ya se han analizado los requisitos y se ha pasado a la fase de Obtención de los Datos. El sector de Desarrollo y el de Investigación son los encargados de analizar los mismos y limpiarlos.

5.3.1.2.2. Acordar la distribución de información

Esta actividad de este SubProceso, correspondiente a los Procesos de Administración de Proyectos, se lleva a cabo en las fases Plan de Proyecto y Validación de Objetivos (segunda vuelta del espiral), y Generación Prueba de Diseño y Construcción del Modelo (quinta vuelta del espiral).

Se ha distribuido la información del proyecto al equipo de trabajo para que cada uno en su puesto y función pueda realizar lo correspondiente para el desarrollo del proyecto.

5.3.1.2.3. Revisar con el responsable la descripción del producto, el equipo de trabajo y el calendario

Esta actividad mantiene a las fases Plan de Proyecto y Validación de Objetivos (segunda vuelta del espiral), y Construcción del Modelo y Evaluación del Modelo.(quinta vuelta del espiral). A continuación se lleva a cabo para el Proyecto en cuestión. Se ha llevado a cabo la reunión correspondiente con el responsable del desarrollo del proyecto para que los requisitos establecidos por el cliente estén bien entendidos y no surjan problemas en próximos periodos de tiempo. Se distribuyo la documentación de los requisitos y el plan de trabajo y se repasaron plazos de tiempos y expectativas.

5.3.1.2.4. Revisar cumplimiento del plan de adquisiciones y capacitación

La actividad que se lleva a cabo a continuación está ligada con las fases de Plan de Proyecto y Validación de Objetivos como así también con la de Evaluación del Modelo perteneciente a la quinta vuelta del Modelo de Ciclo de Vida.

Se ha cumplido el Plan de Capacitación, ya que los involucrados han sido capacitados en lo que al problema resulta. Se han aprendidos términos médicos útiles para poder entender de manera mas rápida y mejor la complejidad del problema. Al mismo tiempo se han estudiado las metodologías existentes para llevar a cabo proyectos de este estilo.

5.3.1.2.5. Recolectar reportes de actividades y mediciones y sugerencias de mejora y productos de trabajo

Esta actividad se realiza en las fases de Plan de Proyecto y Validación de Objetivos correspondiente a la segunda vuelta del Modelo de Ciclo de Vida, a Verificación de Calidad de Datos perteneciente a la tercera vuelta del espiral, a Validación de Datos correspondiente a la cuarta vuelta y a Evaluación del Modelo. A continuación se lleva a cabo para el Proyecto que se está desarrollando:

El conjunto de datos contiene 416 registros de pacientes hepáticos y 167 pacientes sin enfermedad de hígado. El conjunto de datos se recogió desde el norte del país, en provincias como Salta, Jujuy, Formosa y Chaco. Selector es una etiqueta de clase que se utiliza para dividir en grupos (paciente hepático o no). Este conjunto de datos contiene 441 registros de pacientes masculinos y 142 registros de pacientes femeninos. El archivo con los datos se encuentra en formato CSV. Los campos son los siguientes:

1. Edad: Edad del paciente
2. Género El género del paciente
3. TB Bilirrubina Total
4. DB bilirrubina directa
5. AlkPhos fosfatasa alcalina
6. SGPT Alanina aminotransferasa
7. SGOT Aspartato aminotransferasa:
8. TP totales proteínas
9. ALB albúmina
10. Relación A / G cociente albúmina y globulina
11. Campo de selección utilizado para dividir los datos en dos conjuntos (etiquetado por los expertos)

5.3.1.2.6. Registrar costo real del proyecto

No se ven ajustes necesarios a realizar

5.3.1.2.7. Revisar el registro de rastreo basado en los productos de trabajo recolectados

El producto de trabajo recolectado es el Modelo desarrollado mediante un árbol de decisión que aporta conocimiento por las reglas que el mismo posee permitiendo diagnosticar más eficientemente. Este producto se graba en el Reporte de Seguimiento al Cliente.

5.3.1.2.8. Revisar los productos terminados durante el proyecto

Los productos terminados son dos modelos, uno original y otro alternativo el cual se obtiene un árbol de decisión más detallado.

5.3.1.2.9. Recibir y analizar las solicitudes de cambio del cliente

No se han recibido solicitudes de cambio por parte del cliente y se le ha informado de los modelos desarrollados.

5.3.1.2.10. Realizar reuniones con el equipo de trabajo y cliente para reportar avances y tomar acuerdos

Se realizaron reuniones con el equipo de trabajo para verificar los avances realizados y los productos terminados y corroborar si se cumplió con el Plan del Proyecto. El mismo fue cumplido.

5.3.1.3. Evaluación y Control

Este SubProceso cuenta con las siguientes actividades: Evaluar el cumplimiento del plan de proyecto y plan de desarrollo, Analizar y controlar los riesgos, y Generar el reporte de seguimiento del proyecto.

5.3.1.3.1. Evaluar el cumplimiento del plan de proyecto y plan de desarrollo

Esta actividad se lleva a cabo luego del desarrollo del Modelo, por lo tanto las fases a las que mantendrá esta actividad son a las de la última vuelta del espiral, o sea Evaluación de Resultados, Revisión Proceso, Determinación Próximos Pasos, Desarrollo Plan, Plan Supervisión y Mantenimiento, Informe Producto y Revisión Proyecto. Al momento se ha cumplido en gran parte el Plan de Proyecto definido en secciones anteriores. En cuanto a los ciclos, se están llevando a cabo cada uno de los 7 ciclos definidos. Los costos acarreados se estipula que estén dentro de la estimación desarrollada, salvo que se le daría un bono a los involucrados del proyecto dado el cumplimiento de los plazos y si el proyecto finaliza en menos del tiempo esperado. El equipo de trabajo inicial detallado en secciones anteriores no se ha modificado. Se han realizado mínimos ajustes en el Plan de Proyecto. En primer lugar, como se ha mencionado, se ha incrementado en

pequeña proporción los costos acarreados al proyecto, ya que como premio e incentivo al haber avanzado en tiempos deseados se le daría a los involucrados un bono mínimo.

Por otro lado, como ya se ha dicho, los tiempos estipulados para el desarrollo total del proyecto se ha respetado, más allá de que los tiempos de las tareas han sido finalmente menores a los pensados.

Teniendo en cuenta el Plan de Desarrollo descrito en secciones anteriores del presente trabajo, se puede decir que el mismo se ha estado cumpliendo teniendo en cuenta diferentes ejes: en primer lugar el proyecto como se ha provisionado ha sido factible de desarrollar, en segundo lugar las metas se han alcanzado y las técnicas a utilizar se han utilizado de manera exitosa, y en tercer lugar los recursos han sido suficientes para el desarrollo del proyecto.

5.3.1.3.2. *Analizar y controlar los riesgos*

Dicha actividad se realiza en las últimas fases del Proyecto, específicamente en la de Desarrollo Plan, Plan Supervisión y Mantenimiento, y Revisión de Proyecto. A continuación se lleva a cabo para el Proyecto de E.I. actual. Teniendo en cuenta los posibles riesgos identificados en secciones anteriores, y las contingencias para cada riesgo, se presenta una descripción de cada riesgo teniendo en cuenta su aparición o no en el proyecto (tabla 5.3).

Riesgo	Aparición
Aparición de cambios en los requerimientos por parte del cliente.	Los requerimientos no se han modificado
El riesgo organizacional que podrían llegar a presentarse es la ida de algún miembro del equipo involucrado en este proyecto.	El equipo de desarrollo y los involucrados en el proyecto han sido los mismo que se han detallado
Los riesgos técnicos encontrados son que los recursos materiales de la organización sufran algún siniestro.	No se ha presentado ningún siniestro
Datos no sean lo suficientemente buenos como para poder realizar este proyecto.	Los datos han sido suficientes y validos como para poder llevar a cabo el proyecto.

Table 5.3. Plan de Contingencia

En adición a esos riesgos identificados desde un primer momento, el grupo ha identificado un nuevo riesgo que puede llegar a ocurrir. El mismo se detalla a continuación con su respectiva acción:

Riesgo	Acción
El cliente no entiende el léxico técnico utilizado.	Se desarrollara un manual con el léxico técnico utilizado a fin de que el cliente entienda de lo que se está hablando

Table 5.4. Plan de Contingencia para Nuevo Riesgo

Teniendo en cuenta el Plan de riesgos detallado en secciones anteriores del presente trabajo, y el nuevo riesgo identificado en el transcurso de este proyecto de explotación de información, se ha ajustado el plan de riesgos, teniendo los siguientes riesgos y su acción si este ocurriese en la tabla 5.5.

5.3.1.3.3. Generar el reporte de seguimiento del proyecto

La siguiente actividad se realiza en las ultimas 4 fases del Modelo de Ciclo de Vida utilizado en este Proyecto de Explotación de Información. Estas son: Desarrollo Plan, Plan Supervisión y Mantenimiento, Informe Producto, y Revisión Proyecto.

Teniendo en cuenta lo desarrollado en este trabajo y las definiciones primarias sobre el mismo, el equipo de trabajo no ha identificado cosas o acciones que puedan llegar a cambiar en el futuro.

En base a los objetivos detallados de este trabajo y a lo desarrollado, se puede sentenciar que este trabajo ayudara al Ministerio de Salud de la República Argentina, y por ende, a todas las entidades de salud del país a identificar de manera más rápida y eficiente cuando un paciente con síntomas similares posee alguna enfermedad hepática y cuando no. Esto ayudara en el largo plazo, colaborando directamente con la atención y aportando conocimiento para que las decisiones sean oportunas y correctas, logrando ganar tiempo para atacar la enfermedad que puede llegar a tener el paciente.

5.3.1.4. Cierre /Entrega

Este SubProceso cuenta con las siguientes actividades desarrolladas mas abajo: Formalizar la terminación del proyecto o ciclo, Llevar a cabo el cierre del contrato con subcontratistas, Generar el reporte de mediciones y sugerencias de mejora, y Planear la entrega.

5.3.4.1. Formalizar la terminación del proyecto o ciclo

La primera actividad del último SubProceso perteneciente a los Procesos de Administración de Proyectos mantiene a las fases de Desarrollo Plan, Plan Supervisión y Mantenimiento, Informe Producto, y Revisión Proyecto.

Riesgo	Acción
Aparición de cambios en los requerimientos por parte del cliente.	Entrevista más profunda para no sufrir cambios de nuevo. Cambios en los tiempos estipulados si el requerimiento fue cambiado por cliente.
El riesgo organizacional que podría llegar a presentarse es la ida de algún miembro del equipo involucrado en este proyecto.	Capacitación a más de un recurso humano para que el cambio no sea atroz
Los riesgos técnicos encontrados son que los recursos materiales de la organización sufran algún siniestro.	Almacenamiento de recursos materiales adicionales para suplir a los dañados
Datos no sean lo suficientemente buenos como para poder realizar este proyecto.	Búsqueda de varias fuentes de datos para obtener alternativas y gran variedad.
El cliente no entiende el léxico técnico utilizado.	Se desarrollara un manual con el léxico técnico utilizado a fin de que el cliente entienda de lo que se está hablando.

Table 5.5. Plan de Contingencia Final

Teniendo en cuenta la finalización del proyecto se da paso al cierre y a la entrega del producto. Por ende, se le ha comunicado formalmente en el día de la fecha al cliente la finalización del proyecto, comunicando que lo requerido ha sido logrado.

Además, se ha documentado la finalización de dicho proyecto dentro de la Consultora y se ha notificado a todos los sectores de la Organización involucrados en el proyecto.

5.3.1.4.2. Llevar a cabo el cierre del contrato con subcontratistas

Esta actividad, al igual que la anterior, se lleva a cabo en las fases Desarrollo Plan, Plan Supervisión y Mantenimiento, Informe Producto, y Revisión Proyecto. A continuación se desarrolla para el Proyecto actual:

Dada la finalización del proyecto en cuestión, se ha llevado a cabo el cierre de cada contrato de la Organización con el cliente, y, además, se ha pasado el proyecto a un historial de la Consultora,

para que los distintos sectores estén al tanto del cumplimiento y finalización del mismo, y cada uno lleve a cabo las tareas que le compete.

5.3.1.4.3. Generar el reporte de mediciones y sugerencias de mejora

Tal como las dos primeras actividades del actual SubProceso, esta actividad se lleva a cabo en las últimas cuatro fases del Modelo de Ciclo de Vida propuesto en el capítulo anterior.

Se ha llevado a cabo una entrevista con cada involucrado dentro del proyecto para poder conocer sus experiencias a lo largo del mismo. En las mismas se han manifestado aspectos positivos en cuanto al desarrollo de este proyecto, ya que ha permitido tener contacto con un cliente muy importante, el cual permitirá el desarrollo de nuevos proyectos en distintos ámbitos (tener contacto con el Ministerio de Salud de la República Argentina permite que él mismo u otro Ministerio se interese en llevar a cabo proyectos del mismo índole), al mismo tiempo ha permitido a los involucrados en el desarrollo ganar experiencia en este tipo de proyectos, logrando conocimiento más fuerte y sustentable en tecnologías de explotación de información, creciendo también como analistas ya que el trabajo solicitado es algo nuevo y distinto a los temas que se venían desarrollando.

En cuanto al cliente, se ha llevado a cabo una entrevista con uno de los usuarios del proyecto y ha manifestado su satisfacción con el mismo, manifestando que el mismo producirá diagnósticos más específicos y certeros, permitiendo dar conocimiento a los médicos del país.

El feedback con el cliente ha sido muy bueno y de forma justa, ya que las entrevistas tanto como para especificar los requisitos del proyecto, como para ir mostrando los avances y finalización del mismo se ha desarrollado en buenos términos, se han realizado charlas amenas, dejando las cosas claras y sin malos entendidos. Esta experiencia con el Ministerio de Salud de la República Argentina le abrirá nuevas puertas a la consultora, permitiendo conseguir un mayor número de clientes ya que se ha dado una publicidad de la misma y, el cumplimiento de los términos y finalidades del proyecto, permitirá que los futuros clientes confíen en el equipo de trabajo de Consulting.

El proceso completo de este proyecto ha transcurrido en los cauces normales y esperados. La consultora en su conjunto, incluyendo los sectores involucrados (no solo el de Informática), ha llevado a cabo un excelente trabajo, desde la Administración, la gestión, el desarrollo a la contabilidad. Se ha desarrollado de manera exitosa el proyecto, cumpliendo los requisitos solicitados y expuestos por el cliente, llevando a cabo una documentación del mismo el cual servirá de base y de experiencia para nuevos proyectos que se presenten.

5.3.1.4.4. Planear la entrega

Finalizando los Procesos de Administración de Proyectos y con el Producto ya terminado, a nivel administrativo esta tarea se realiza bajo las siguientes cuatro fases del Modelo de Ciclo de Vida: Desarrollo Plan, Plan Supervisión y Mantenimiento, Informe Producto, y Revisión Proyecto.

La entrega se llevará a cabo de la siguiente manera: se le entregará al cliente en mano toda la documentación correspondiente al proyecto, el programa en sí en un CD indicando por escrito los pasos a seguir para que cualquier usuario que ellos deseen puedan instalar el aplicativo y usarlo de manera sencilla. Además se firmara un contrato el cual indica el cierre del proyecto y la aceptación del cliente, comprometiéndose la consultora a darle soporte y mantenimiento al Ministerio.

5.3.2. Procesos de Desarrollo de Proyectos

En esta sección se lleva a cabo el desarrollo del Proyecto de Explotación de Información teniendo en cuenta los SubProcesos de Desarrollo.

5.3.2.1. Entendimiento del Negocio

En este SubProceso se llevan a cabo la siguiente tarea: Determinar las metas del proyecto de Explotación de Información.

5.3.2.1.1. Determinar las metas del proyecto de Explotación de Información

Esta actividad cubre y se realiza en las fases tanto de la primer vuelta como de la segunda correspondiente al Modelo de Ciclo de Vida del Proyecto: Objetivos del Negocio, Evaluación de la Situación, Objetivos del Proyecto de Explotación de Información, Plan del Proyecto y Validación de Objetivos. A continuación se realiza para el Proyecto actual:

A partir de los interrogantes mencionados, tales como Que son las enfermedades hepáticas?, o Cuales son ellas? o como así también Cuáles son los síntomas más significativas para diagnosticar alguna enfermedad del hígado?, este Proyecto de Explotación de Información tiene como metas principal brindar a partir de su funcionamiento herramientas y conocimiento a las empresas que se encarguen de la salud para que puedan tomar decisiones más acertadas y certeras, disminuyendo el riesgo y la incertidumbre y que los doctores o encargados de brindarle al paciente una respuesta, puedan diagnosticar a los mismos de forma más rápida y concreta. Técnicamente hablando, el Proyecto de Explotación de Información desarrollado en este trabajo, tiene como meta lograr a partir de los datos obtenidos y limpios, lograr saber bajo ciertas reglas si un paciente que se atiende en alguna entidad de salud posee o no alguna enfermedad hepática.

Teniendo en cuenta esta principal meta, se tiene como problema la intención de lograr por medio de este trabajo, llevar a cabo un descubrimiento de reglas de comportamiento. Así, se puede determinar

con qué valores y atributos un paciente que se atiende o manifieste síntomas de alguna enfermedad, realmente está enfermo o si no se detecta alguna enfermedad hepática.

Los modelos a presentar deben tener cualidades para poder finalmente ser aceptados. En primer lugar, el/los modelo/s deben ser comprensibles así el usuario final puede entenderlo y mediante ellos puede rápidamente tomar una decisión con el conocimiento que el modelo y el Proyecto a realizar dará. En segundo lugar, este modelo debe brindar resultados concretos y que mediante la devolución que él hará, el encargado de utilizar el mismo pueda fácilmente tomar una decisión y diagnosticar de manera más eficaz y con certidumbre. En tercer lugar, los atributos y reglas de comportamiento deben ser determinantes y no generarle duda al usuario que utilice la herramienta. Así, mediante todo esto descripto, el agente puede diagnosticar al paciente de mejor manera, sin malos entendidos ni incertidumbre que cause malos diagnósticos, pérdida de tiempo y mayores costos. Se ejecutarán varias pruebas de la solución propuesta por este proyecto en distintas máquinas para evaluar si el rendimiento es similar en cada una de ellas. Además se variaran la cantidad de valores e información para ver la variación de tiempo de ejecución. Se espera que si el tiempo varía sea lo menos posible. Mediante la utilización de benchmarks se intentará evaluar el rendimiento de cada uno de los procesos. Tal como se ha explicado, el éxito de este Proyecto de Explotación de Información a realizar, estará dado si los resultados que el modelo brinde aporten conocimiento suficiente y colaboren con la toma de decisiones de los encargados de diagnosticar a pacientes hepáticos. Además, técnicamente hablando, el éxito se dará si cualquier agente médico puede comprender fácil y rápidamente las reglas de comportamiento y no queda margen de duda o permita la aparición de creencias mal formadas o malos entendidos.

Teniendo en cuenta el Modelo de Ciclo de Vida propuesto, al finalizar este SubProceso, el cual abarca las fases de la primera y segunda vuelta del espiral, se lleva a cabo en el eje de Planeamiento la Planificación de la Selección de Datos. Luego se realiza un Análisis de Riesgo previo al pasaje al posterior SubProceso. El Análisis de Riesgo llevado a cabo para este SubProceso es el siguiente:

5.3.2.2. Entendimiento de los Datos

Este SubProceso posee las siguientes actividades que se llevan a cabo a continuación: Reunir los datos iniciales, Describir los datos, Explorar los datos, y Verificar la calidad de los datos.

5.3.2.2.1. Reunir los datos iniciales

Esta primera actividad del presente SubProceso se lleva a cabo en la fase de Recolección de Datos perteneciente a la tercera vuelta del Modelo de Ciclo de Vida en Espiral usado en el Proyecto. A continuación se utiliza para el mismo:

Riesgo	Acción
Aparición de cambios en los requerimientos por parte del cliente.	Entrevista más profunda para no sufrir cambios de nuevo. Cambios en los tiempos estipulados si el requerimiento fue cambiado por cliente.
El riesgo organizacional que podrían llegar a presentarse es la ida de algún miembro del equipo involucrado en este proyecto.	Capacitación a más de un recurso humano para que el cambio no sea atroz
Los riesgos técnicos encontrados son que los recursos materiales de la organización sufran algún siniestro.	Almacenamiento de recursos materiales adicionales para suplir a los dañados

Table 5.6. Analisis de Riesgo para Entendimiento del Negocio

La información necesaria para poder llevar a cabo este Proyecto de Explotación de Información y que sirva y sea útil para el Ministerio de Salud de la Republica Argentina y por ende a todos los entes de salud del país, se necesitan valores para atributos que han sido identificado por los expertos para poder determinar mediante esos factores si un paciente que se presenta a una entidad de salud del país posee o no alguna enfermedad hepática. Los factores o información necesaria para poder llevar a cabo este proyecto son:

- * Edad: Edad del paciente
- * Género El género del paciente
- * TB Bilirrubina Total
- * DB bilirrubina directa
- * AlkPhos fosfatasa alcalina
- * SGPT Alanina aminotransferasa
- * SGOT Aspartato aminotransferasa:
- * TP totales proteínas
- * ALB albúmina
- * Relación A / G cociente albúmina y globulina

Mediante distintos valores y experiencias tomadas, se puede determinar si se encuentra una enfermedad o no. Actualmente, y gracias al trabajo de los investigadores, se han recogido registros de pacientes de provincias del Norte de Argentina que se han presentado en un Hospital o Clínica con la intención de que lo diagnostiquen. Se han tomado más de 550 casos, con variaciones en los

valores de los atributos identificados anteriormente. La información se encuentra disponible para poder llevar a cabo este Proyecto.

Para llevar a cabo la selección de la información dentro de este Proyecto de Explotación de Información, se han detectado una serie de criterios. En primer lugar se desean datos únicos y que no se encuentren registros duplicados que entorpezcan el accionar del Proyecto. En segundo lugar se seleccionarán variedad de información, valores que sean lo suficientemente variado para poder obtener distinto tipos de reglas y que permitan lograr un diagnostico con mayor precisión y certeza. Además se desea que la información sea real, que no se encuentre adulterada y ayude al diagnostico de pacientes dentro de nuestro país. Por último es necesario que todos los registros posean todos los valores y no se produzca la aparición de datos nulos.

Para poder llevar a cabo el modelo será necesario cambiar el formato del archivo obtenido con la información disponible con los 500 casos experimentados en provincias como Salta, Jujuy, Formosa y Chaco. Esta modificación consta que el archivo está en formato CSV (.csv) y para poder realizar el Proyecto de Explotación de Información comentado en este trabajo, es necesario transformarlo a un archivo Excel (.xls). Además es necesaria la inclusión de un registro delante de todo, el cual será el titulo de cada uno de los campos.

A esta información se puede acceder de forma gratuita, la explicación de que es cada columna se encuentra misma en la página web. Dado que son más de 550 registros con diferentes resultados, se considera que la información disponible es suficiente para llevar a cabo este Proyecto.

5.3.2.2.2. Describir los datos

Esta segunda actividad se utiliza en la fase de Descripción de Datos. Siguiendo con el desarrollo del Proyecto de E.I. se lleva a cabo la actividad mencionada:

Los datos fueron adquiridos por el personal de investigación, detectando la necesidad del proyecto y evaluando que datos son necesarios para poder llevar a cabo este Proyecto. Así obtuvieron una base de datos de pacientes hepáticos de provincias del Norte del País, los mismos datos se consiguieron de forma gratuita y abarcan la cantidad deseada para poder realizar este Proyecto. Como ya se mencionó anteriormente los datos necesarios para la realización de este Data Mining fueron: Edad, Género, TB Bilirrubina Total, DB bilirrubina directa, AlkPhos fosfatasa alcalina, SGPT Alanina aminotransferasa, SGOT Aspartato aminotransferasa, TP totales proteínas, ALB albúmina, Relación A / G cociente albúmina y globulina y una columna extra que indique el resultado (posee o no enfermedad hepática).

Los datos que se obtuvieron a partir de la investigación, no poseen entrada de texto libre. Los atributos se encuentran accesibles y son de fácil lectura, sin generar dudas o malos entendidos. Los rangos obtenidos de los valores de los atributos de los datos obtenidos son los especificados, por lo

cual se los consideran validos. En la tabla 5.7 se ve la definición de los campos y en la tabla 5.8 se presentan los rangos de los valores de los datos obtenidos.

Se espera que con esta variación de valores se obtengan reglas que sean significativas y aporten de gran manera al diagnostico de enfermedades hepáticas.

Finalmente se ha entrevistado al experto en el tema y ha aprobado el volumen y calidad de casos que se han tomado, indicando que son los necesarios para poder llevar a cabo el Proyecto.

5.3.2.2.3. Explorar los datos

Siguiendo en el SubProceso Entendimiento de los Datos, esta actividad se realiza en la fase Exploración de los Datos perteneciente a la tercera vuelta del Modelo de Ciclo de Vida.

Se han analizado las propiedades de los atributos del volumen de datos obtenidos y se ha decretado que todos son interesantes ya que todos son importantes para poder diagnosticar de manera eficiente. Por ejemplo se detecta que la Edad es un factor importante, como así también la relación A/G cociente albumina y globulina y la cantidad de proteínas que el paciente posee.

Considerando y evaluando la información y los resultados en los reportes de descripción de datos no se han detectado fallas ni necesidad de modificación o arreglo.

Mediante estos datos, se han identificado una serie de hipotesis que se definen a continuación:

Hipotesis 1: Las personas más jóvenes tienden menos a contraer enfermedades hepáticas

Meta 1: Mediante esta hipótesis se tomarán registros de prueba con edades significativas para ratificar o refutar la hipótesis en cuestión.

Hipotesis 2: Existe una relación entre el sexo del paciente a diagnosticar con el diagnostico finalmente brindado a él.

Meta 2: Se tomaran registros de prueba, teniendo misma cantidad de mujeres que de hombres. Así, mediante descubrimiento de reglas de comportamiento se evaluará quien es mas propenso o si no existe relacion entre el sexo y el diagnóstico.

Hipotesis 3: La bilirrubina total tiende a ser menos a 10, salvo en casos extremos y ese cambio impacta en el diagnostico final.

Meta 3: Se evaluará al final casos con más y menos de 10 de bilirrubina total para saber si influye significativamente o no en el diagnostico final.

Las acciones a desarrollar teniendo en cuentas las hipótesis presentadas anteriormente, son las de analizar luego de descubrir las reglas de comportamiento esas hipótesis.

La meta final de este Proyecto de Explotación de Información como se ha dicho a lo largo del trabajo es identificar por medio de algunos atributos si el paciente posee o no enfermedad hepática.

Campo	Tipo de Dato
-------	--------------

Edad	Numérico (4 a 90)
Genero	String (“Female” o “Male”)
TB Bilirrubina Total	Numérico
DB bilirrubina directa	Numérico
AlkPhos fosfatasa alcalina	Numérico
SGPT Alanina aminotransferasa	Numérico
SGOT Aspartato aminotransferasa	Numérico
TP totales proteínas	Numérico
ALB albúmina	Numérico
Relación A / G cociente albúmina y globulina	Numérico
Diagnostico Final	Numérico (1 –posee, 2- no posee)

Table 5.7. Definición de datos

5.3.2.2.4. Verificar la calidad de los datos

La última actividad de este SubProceso se realiza bajo la fase Verificación Calidad de Datos, siendo desarrollada a continuación:

Teniendo en cuenta los datos obtenidos por el equipo de investigación no se han encontrado valores especiales dentro de los mismos. Se ha identificado que los atributos y los valores de los mismos encajan entre sí. Al mismo tiempo se identificó que no se encuentran valores faltantes o datos en blancos, cada registro posee todos los valores, lo cual ayuda al desarrollo del Data Mining.

La escritura de los valores de los campos es similar en todos, salvo el campo Sexo (Genero) que es de tipo String, a diferencia de todos los demás que son numéricos. Algunos Numéricos como el diagnostico final y la edad son enteros, mientras que el resto puede ser con decimales.

El archivo se encuentra en formato CSV, teniendo como delimitador el carácter coma (“,”). No se han encontrado ruidos en los atributos de este lote de datos.

Campo	Valor	
	Min	Valor Max
Edad	4	90
Genero	Female	Male
TB Bilirrubina Total	0.4	75
DB bilirrubina	0.1	19.7

directa		
AlkPhos fosfatasa alcalina	63	2110
SGPT Alanina aminotransferasa	10	2000
SGOT Aspartato aminotransferasa	10	4929
TP totales proteínas	2.7	9.6
ALB albúmina	0.9	5.5
Relación A / G cociente albúmina y globulina	0.3	2.8
Diagnostico Final	1	2

Table 5.8. Posibles valores para datos

Al finalizar este SubProceso y según el Modelo de Ciclo de Vida en Espiral Propuesto, se realiza la Planificación de la Preparación de los Datos. Por otro lado, en el eje de Evaluación de Alternativas e Identificación y Resolución de Riesgos, se desarrollo un Análisis de Riesgo previo al pasaje a la siguiente vuelta del espiral. El Análisis de Riesgo desarrollado es el siguiente:

Riesgo	Acción
Datos no sean lo suficientemente buenos como para poder realizar este proyecto.	Búsqueda de varias fuentes de datos para obtener alternativas y gran variedad.

Table 5.9. Analisis de Riesgo para Entendimiento de los Datos

5.3.2.3. Preparación de los Datos

Este SubProcesos cuenta con las siguientes actividades a ser desarrolladas a continuación: Tareas Preparatorias, Seleccionar los datos, Limpiar los datos, Construir los datos, Integrar los datos, y Formatear los datos.

5.3.2.3.1. Tareas Preparatorias

Esta primera actividad se realiza en la fase Selección de Datos, la primera actividad perteneciente a la cuarta vuelta del Modelo de Ciclo de Vida. A continuación se desarrolla esta actividad para el Proyecto que se usa como Prueba de Concepto:

Se han recogido alrededor de 550 casos aproximadamente de pacientes de Salta, Jujuy, Formosa y Chaco, a los cuales a algunos se le diagnostico que posee enfermedad hepática y a otros no. Los mismos cuentan con los mismos atributos, variando los valores.

Este dataset contiene 416 registros de pacientes hepáticos y 167 pacientes que no se les ha diagnosticado enfermedad. El último campo es una etiqueta que se utiliza para dividir en grupos (paciente hepático o no). Este conjunto de datos contiene 441 registros de pacientes masculinos y 142 registros de pacientes femeninos.

5.3.2.3.2. *Seleccionar los datos*

Al igual que la actividad anterior, esta también se realiza sobre la fase Selección de Datos y se lleva a cabo a continuación. Los posibles valores para los datos se ven en la tabla 5.10

El grupo de investigación y desarrollo ha decidido que no es necesario seleccionar o agregar datos adicionales a los que ya se cuenta. Así, se ha detectado la necesidad de contar con los datos recogidos anteriormente, sin tener que agregar o quitar alguno de ellos.

Campo	Tipo de Dato
Edad	Numérico (4 a 90)
Genero	String ("Female" o "Male")
TB Bilirrubina Total	Numérico
DB bilirrubina directa	Numérico
AlkPhos fosfatasa alcalina	Numérico
SGPT Alanina aminotransferasa	Numérico
SGOT Aspartato aminotransferasa	Numérico
TP totales proteínas	Numérico
ALB albúmina	Numérico
Relación A / G cociente albúmina y globulina	Numérico
Diagnostico Final	Numérico (1 –posee, 2- no posee)

Table 5.10. Posibles valores para datos

Para el muestreo de los datos se utilizaran las siguientes técnicas:

- Técnica de Muestreo Aleatorio Simple: Dado que se tienen 583 registros, 441 de los cuales pertenecen al sector masculino y el restante al femenino, se extraerá una muestra de 30 pacientes enumerando a los pacientes del 1 al 583, luego sorteando 30 números entre los 583 y la muestra estará dada por los 30 registros que corresponden a los números obtenidos.

- Técnica de Muestreo Aleatorio Sistemático: Teniendo la misma cantidad de registros descripta en la muestra anterior, y queriendo extraer la misma cantidad de muestra, se enumeran los pacientes del 1 al 583, se calcula el intervalo constante entre cada individuo ($583/30=19.43$), se sortean números del 1 al 19. Dependiendo del número sorteado, los demás 29 pacientes serán en el intervalo del número sorteado. Ejemplo: Sale sorteado el paciente 15, por lo tanto para la muestra se tomara el paciente 15, 30, 45...120, etc.

5.3.2.3.3. Limpiar los datos

Limpieza de Datos es la fase en la que se ejecuta esta actividad perteneciente al SubProceso Preparación de los Datos. Siguiendo con el desarrollo del Proyecto, se lleva a cabo la actividad mencionada:

Como se ha mencionado en la sección anterior, para este Proyecto de Explotación de Información, no se han detectado ruidos posibles que afecten el cauce normal del mismo. Así, no resulta necesario tener que corregir, remover o ignorar los mismos.

No se han detectado valores especiales ni en blanco en los datos recogidos por el personal. Lo único que se ha modificado de los mismos es cambiar los valores 1 y 2 del último atributo (“Diagnostico Final”) que indica si un paciente posee o no enfermedades hepáticas por “Positivo” (para indicar la presencia de enfermedad) o “Negativo” (para indicar que no se ha detectado, por medios de los síntomas indicados, la presencia de enfermedad). Además, resulto necesario modificar los valores que poseía el atributo sexo, para poder llevar a cabo correctamente la ejecución del algoritmo (Male por 1 y Female por 2).

5.3.2.3.4. Construir los datos

Una vez limpiado los datos, la actividad a llevar a cabo es la de construirlos. Esta actividad se realiza bajo la fase Construcción de Datos y se desarrolla a continuación.

No se considera necesario Reconsiderar los criterios de selección de los datos basándose en las experiencias de construcción de datos.

Para el entendimiento del modelo a desarrollar, se ha decidido normalizar el ultimo campo que indica si el paciente está enfermo o no, cambiando los valores 1 y 2 por Positivo o Negativo. Así es más claro para el usuario al momento de dar su diagnostico.

Como ya se ha mencionado anteriormente, no se ha detectado la necesidad de agregar nueva información a la que ya se posee.

Se ha agregado al archivo el titulo a cada campo permitiendo identificar en un futuro a que campo pertenece cada valor. Además como se ha dicho anteriormente, se cambiaron los valores del campo

“Diagnostico Final”, permitiendo una mejor comprensión del resultado final y los valores de Sexo para poder ejecutar correctamente el algoritmo seleccionado.

5.3.2.3.5. Integrar los datos

Esta actividad se lleva a cabo en la fase Integración de Datos, siguiendo situados en la cuarta vuelta del Modelo de Ciclo de Vida.

Los datos recogidos por el personal ya se encuentran integrados. Los mismos permiten una rápida comprensión y el agente medico va a poder fácilmente diagnosticar a un paciente, sin que haya cosas libradas al azar. No se detecto la necesidad de reconsiderar los criterios de selección de datos ya que los mismos, como se ha dicho anteriormente, se encuentran integrados y abarcan la necesidad del Proyecto.

5.3.2.3.6. Formatear los datos

La última actividad de este SubProceso es la de formatear los datos, y la misma se realiza en la fase Formateo de Datos del Modelo de Ciclo de Vida. A continuación se lleva a cabo para el Proyecto actual. Dado que la herramienta seleccionada para llevar a cabo este Proyecto de Explotación de Información (Tanagra), solicita que el archivo sea un Excel y con header, se ha cambiado la extensión del archivo con los datos, pasando de CSV a XLS. Además, dado que se necesita header, se le ha agregado un registro al principio, indicando en cada columna el nombre de cada atributo. Así, es más fácil identificar qué valor se está evaluando. No se cree necesario y/o conveniente reordenar los registros del archivo de datos. Como se menciono anteriormente, fue necesario el formateo del campo “Diagnostico Final” pasando de valores 1 y 2 a tener “Positivo” y “Negativo” respectivamente. Esto se realizo ya que quede más claro al ver el modelo si el paciente posee o no posee enfermedad hepática. Esto permite mejor usabilidad y comprensión. Además, resultado necesario modificar los valores que poseía el atributo sexo, para poder llevar a cabo correctamente la ejecución del algoritmo (Male por 1 y Female por 2). Por lo tanto, la definición de los datos ha quedado definida en la tabla 5.11.

No se considera necesario reconsiderar la selección de datos en base a experiencias de limpieza de datos.

Cuando se terminan estas actividades pertenecientes a las fases de la cuarta vuelta del Modelo de Ciclo de Vida en Espiral se lleva a cabo la Planificación del Modelo.

Campo	Tipo de Dato
Edad	Numérico (4 a 90)
Genero	Numérico (1 o 2)
TB Bilirrubina Total	Numérico
DB bilirrubina directa	Numérico

AlkPhos fosfatasa alcalina	Numérico
SGPT Alanina aminotransferasa	Numérico
SGOT Aspartato aminotransferasa	Numérico
TP totales proteínas	Numérico
ALB	Numérico
Relación A / G cociente albúmina y globulina	Numérico
Diagnostico Final	String (“Positivo” o “Negativo”)

Table 5.11. Posibles valores para Datos

Por otro lado se desarrolla otro Análisis de Riesgo para identificar los mismos y tratarlos. El Análisis de Riesgo para este SubProceso es el siguiente (tabla 5.12):

Riesgo	Acción
Datos no sean lo suficientemente buenos como para poder realizar este proyecto.	Búsqueda de varias fuentes de datos para obtener alternativas y gran variedad.

Table 5.12. Analisis de Riesgo para Preparación de los Datos

5.3.2.4. Modelado

El cuarto SubProceso correspondiente a los Procesos de Desarrollo de Proyectos posee las siguientes actividades dentro: Seleccionar la técnica de modelado, Generar el diseño de test, Construir el modelo, y Evaluar el modelo.

5.3.2.4.1. Seleccionar la técnica de modelado

Esta primera actividad correspondiente al Modelado, se realiza en la fase Selección Técnica Modelado correspondiente a la quinta vuelta del Modelo de Ciclo de Vida en Espiral. Esta quinta vuelta, tal como se menciona en el análisis del Modelo Propuesto en el capítulo 4, itera tantas veces hasta generar el Modelo correcto según los requerimientos que se han definido. A continuación se lleva a cabo esta actividad para el Proyecto que se está desarrollando:

Teniendo en cuenta que se ha elegido a Tanagra como herramienta seleccionada para llevar a cabo este Proyecto de Explotación de Información, y, que en base al problema y requerimiento por parte del cliente de generar un modelo que permita, mediante valores para ciertos atributos identificados, diagnosticar con mayor precisión y eficacia si un paciente se encuentra enfermo o no, se ha elegido la técnica Descubrimiento de Reglas de Comportamiento, utilizando algoritmos TDIDT. Esta

técnica aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema. Así, por medio de esta técnica y la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributos clase, se podrá ver por medio de los valores de los distintos atributos a evaluar un paciente posee o no posee enfermedad hepática.

La técnica seleccionada para el modelado es, como se ha mencionado anteriormente la de Descubrimiento de Reglas de Comportamiento. Mediante la utilización del algoritmo de inducción TDIDT C4.5 se podrán visualizar las condiciones que se tendrán para que a un paciente bajo esos síntomas si posee o no enfermedad hepática. Este algoritmo posee como precondition la presencia de al menos un atributo discreto en el Target, y al menos uno o mas atributos discreto/continuos en la entrada.

Comparando esta asunción y condiciones del algoritmo a utilizar en el modelo con la descripción de los datos se puede afirmar que la utilización de este algoritmo es viable ya que los valores de entrada serán continuos y el target (“Diagnostico Final”) será discreto (“Positivo”, “Negativo”).

5.3.2.4.2. Generar el diseño de test

La fase Generación Prueba Diseño es donde se realiza esta actividad perteneciente al SubProceso Modelado. Siguiendo con el desarrollo del Proyecto de E.I. que es usado para Prueba de Diseño del presente Trabajo Final de Licenciatura, se realiza la actividad mencionada. Tal como se ha mencionado en secciones anteriores, este Proyecto de Explotación de Información tiene como meta principal brindar, a partir de su funcionamiento, herramientas y conocimiento a las empresas que se encarguen de la salud para que puedan tomar decisiones más acertadas y certeras, disminuyendo el riesgo y la incertidumbre y que los doctores o encargados de brindarle al paciente una respuesta, puedan diagnosticar a los mismos de forma más rápida y concreta.

Técnicamente hablando, el Proyecto de Explotación de Información desarrollado en este trabajo, tiene como meta lograr a partir de los datos obtenidos y limpios, lograr saber bajo ciertas reglas si un paciente que se atiende en alguna entidad de salud posee o no alguna enfermedad hepatica. Mediante la utilización de la técnica anteriormente mencionada y la aplicación del algoritmo de inducción TDIDT C4.5 se puede afirmar que se cumpliría con la meta del Proyecto de Explotación de Información ya que podrá mostrar para diferentes valores y combinaciones de los atributos si el paciente da Positivo o Negativo en su estudio. Los pasos a seguir serán los siguientes:

- Se abrirá el archivo CSV con los datos seleccionados para este Proyecto en un Excel.
- Se dividirá cada columna utilizando como delimitador la “,” y se guardara el archivo con extensión .xls (Excel 1997-2003)
- Se agregará un titulo para cada columna.

- Se modificará el valor de la columna “Diagnostico_Final” (1 por Positivo, 2 por Negativo).
- Se modificará el valor Sexo para que se pueda ejecutar el algoritmo TDIDT C4.5 (Male por 1, Female por 2)
- Se ejecutara la herramienta Tanagra
- Se tomaran como parámetros de Input las 10 primeras columnas y como Target la columna final (Diagnostico Final)
- Se tomara el algoritmo de inducción TDIDT C4.5 y se ejecutará.

Al final de este proceso se ha armado el archivo Excel que se tomará para llevar a cabo el modelo con los 583 registros tomados de prueba y los 11 atributos.

5.3.2.4.3. Construir el modelo

La tercera actividad de este SubProceso es desarrollada bajo la fase Construcción del Modelo. La misma se lleva a cabo a continuación:

Se han seteado los parámetros iniciales, tomando como input para la ejecución del algoritmo de inducción TDIDT C4.5 los siguientes atributos: Edad, Genero, TB Bilirrubina Total, DB Bilirrubina directa, AlkPhos fosfatasa alcalina, SGPT Alanina aminotransferasa, SGOT Aspartato aminotransferasa, TP totales proteínas, ALB albumina y Relacion A/G cociente albumina y globulina; y como Target el Diagnostico Final (Positivo - Negativo).

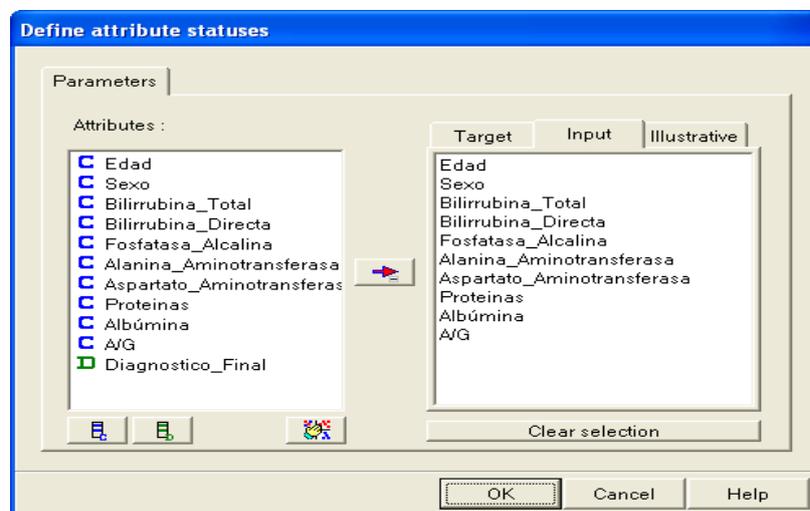


Fig. 5.5. Atributos dentro del Input

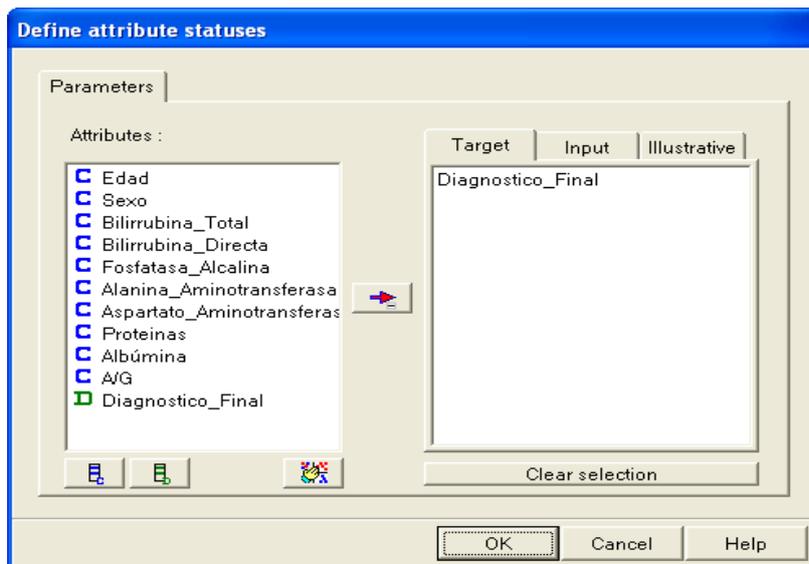


Fig. 5.6. Target

Se han seleccionado esos parámetros ya que se desea determinar mediante esos 10 atributos y sus combinaciones de valores si al paciente se le diagnostica o no la presencia de alguna enfermedad hepática y por ende si necesitan de futuro tratamiento.

Se ha ejecutado la técnica seleccionada por los procesos anteriores sobre los datos seleccionados y limpiados. A continuación se presenta el árbol de decisión (imagen 5.7)

Decision tree

- Bilirrubina_Directa < 1,2500
 - Alanina_Aminotransferasa < 66,5000
 - Edad < 25,5000 then Diagnostico_Final = **Negativo** (61,22 % of 49 examples)
 - Edad >= 25,5000
 - Fosfatasa_Alcaina < 277,0000
 - Edad < 71,0000
 - Edad < 59,0000
 - Bilirrubina_Directa < 0,6500
 - Alanina_Aminotransferasa < 33,0000
 - Aspartato_Aminotransferasa < 21,5000 then Diagnostico_Final = **Positivo** (63,83 % of 47 examples)
 - Aspartato_Aminotransferasa >= 21,5000
 - Edad < 45,5000 then Diagnostico_Final = **Negativo** (62,22 % of 45 examples)
 - Edad >= 45,5000 then Diagnostico_Final = **Positivo** (53,33 % of 30 examples)
 - Alanina_Aminotransferasa >= 33,0000 then Diagnostico_Final = **Positivo** (68,97 % of 58 examples)
 - Bilirrubina_Directa >= 0,6500 then Diagnostico_Final = **Positivo** (77,27 % of 22 examples)
 - Edad >= 59,0000 then Diagnostico_Final = **Negativo** (58,00 % of 50 examples)
 - Edad >= 71,0000 then Diagnostico_Final = **Positivo** (85,00 % of 20 examples)
 - Fosfatasa_Alcaina >= 277,0000 then Diagnostico_Final = **Positivo** (84,09 % of 44 examples)
 - Alanina_Aminotransferasa >= 66,5000 then Diagnostico_Final = **Positivo** (89,39 % of 66 examples)
 - Bilirrubina_Directa >= 1,2500 then Diagnostico_Final = **Positivo** (95,27 % of 148 examples)

Fig.5.7. Árbol de decisión

Como ya se ha dicho antes el valor de la bilirrubina directa parece fundamental a la hora de evaluar al paciente, como así también la edad, la alanina y la fosfatasa. A partir de esos valores se pueden llegar a evaluar otros aspectos para determinar si un paciente posee o no enfermedad hepática. El primer aspecto a evaluar es la bilirrubina directa, ya que un valor mayor a 1,25 ya es síntoma de que el paciente posee enfermedad, sin necesidad de evaluar otros aspectos como la relación A/G o la edad.

5.3.2.4.4. *Evaluar el modelo*

La última actividad de este SubProceso es la de evaluar el modelo, y la misma se lleva a cabo en la fase Evaluación del Modelo del Modelo de Ciclo de Vida.

Teniendo en cuenta los criterios de evaluación del modelo descriptos anteriormente se puede decir que el modelo logro los beneficios esperados.

En primer lugar, el modelo puede ser comprensible por el usuario final, ya que es fácil de leer e interpretar, y no deja lugar a malos entendidos o confusiones que puedan conllevar a un errado diagnóstico.

En segundo lugar este modelo brindó resultados concretos y, mediante la devolución que hace, el encargado de utilizar el mismo pueda fácilmente tomar una decisión y diagnosticar de manera más eficaz y con certidumbre.

En tercer lugar, los atributos y reglas de comportamiento son determinantes y no generan dudas al usuario que utilice la herramienta. Así, mediante todo esto descripto, el agente puede diagnosticar al paciente de mejor manera, sin malos entendidos ni incertidumbre que cause malos diagnósticos, pérdida de tiempo y mayores costos.

Se han cumplido las metas pactadas y descriptas en procesos anteriores sobre este Proyecto de Explotación de Información. Se brindará a partir de su funcionamiento herramientas y conocimiento a las empresas que se encarguen de la salud para que puedan tomar decisiones más acertadas y certeras, disminuyendo el riesgo y la incertidumbre y que los doctores o encargados de brindarle al paciente una respuesta, puedan diagnosticar a los mismos de forma más rápida y concreta. Técnicamente hablando, el Proyecto de Explotación de Información desarrollado en este trabajo logró saber y dar a conocer bajo ciertas reglas si un paciente que se atiende en alguna entidad de salud posee o no alguna enfermedad hepática. Esto produce mayor conocimiento a las organizaciones, mejorando la toma de decisiones.

La información descubierta y el conocimiento brindado a partir de este modelo y este Proyecto es muy útil y novedoso para todas las entidades de salud que deseen diagnosticar con mayor rapidez y eficacia a sus pacientes.

Los resultados obtenidos por medio del modelo se han calificado como confiables. Además, los mismos son lógicos, factibles, son suficientes como para poder llevar a cabo un diagnostico confiable y certero, y no ofenden al sentido común.

Teniendo en cuenta las hipótesis y metas descriptas en secciones anteriores, se ha confirmado que pacientes con menor edad son menos propensos a contraer alguna enfermedad hepática, como así también que existe una relación muy directa entre la edad y el diagnostico final. Por otro lado, teniendo en cuenta la cantidad de bilirrubina detectada, se ha confirmado que una porción menor a 10 determina sin ninguna evaluación más que el paciente posee enfermedad hepática, no así si es mayor a ese valor.

El equipo de trabajo ha ajustado los parámetros establecidos mediante la utilización de la Herramienta Tanagra para Proyectos de Minería de Datos con el fin de presentar un modelo alternativo más en detalle.

Los parámetros se han configurado de la siguiente manera (se paso de 20 de min size of leaves a 10):

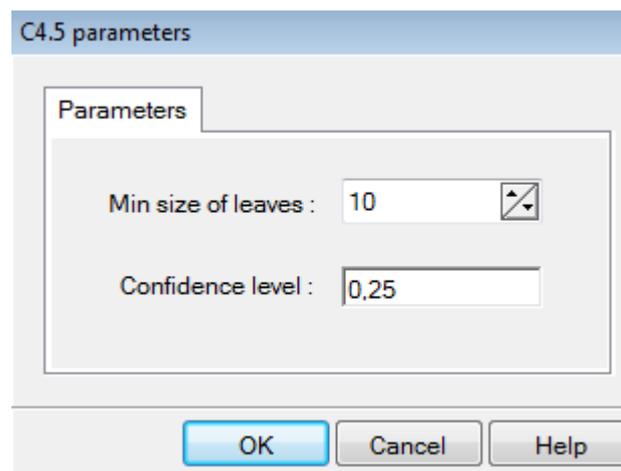


Fig. 5.9. Parámetros Ajustados

Por medio de estos nuevos parámetros se ha ejecutado nuevamente el algoritmo, quedando el modelo de la figura 5.10 para la evaluación del doctor o medico que quiera diagnosticar a sus pacientes.

Por medio de este modelo con parámetros ajustados se puede llevar a cabo un análisis más detallado de los valores de los distintos aspectos tomados para evaluar al paciente. Así, es más seguro el diagnostico y se pueden tomar mejores decisiones.

Cuando se termina este SubProceso de Modelado, se lleva a cabo el desarrollo de un Prototipo para evaluar si el Modelo cumple con los requisitos establecidos. Como en este trabajo de Explotación de Información desarrollado se ha especificado que los mismos han sido cumplido se pasa a la siguiente fase, donde se realiza la Planificación de Evaluación y Despliegue, realizando

posteriormente el ultimo Análisis de Riesgos del Ciclo de Vida del Proyecto. El/Los Análisis de Riesgos desarrollado/s en este SubProceso son:

| Riesgo | Acción |
|--|---|
| Los modelos definidos pueden ser distintos a los que el cliente esperó. | Se desarrolla un Prototipo para que el Cliente evalúe el Modelo. Se lleva a cabo una reconsideración de las técnicas seleccionadas y utilización de las mismas. |
| Requisitos malentendidos del proyecto resultaron en la selección de la herramienta de modelado equivocada. | Se desarrolla entrevista con el Cliente para despejar dudas y que se construya el modelo acorde a sus pretensiones y las del negocio. |
| El cliente no entiende el léxico técnico utilizado. | Se desarrollara un manual con el léxico técnico utilizado a fin de que el cliente entienda de lo que se está hablando. |

Table 5.13. Analisis de Riesgo para Modelado

5.3.2.5. Evaluación

Este quinto SubProceso perteneciente a los Procesos de Desarrollo de Proyectos cuenta con las siguientes actividades: Evaluar resultados, Revisar el proceso, y Determinar próximos pasos.

5.3.2.5.1. Evaluar resultados

Como primer actividad de este SubProceso se obtiene evaluar los resultados. Esta actividad se lleva a cabo en la fase Evaluación de Resultados de la última vuelta del Modelo de Ciclo de Vida en Espiral.

Los resultados dados por este Proyecto de Explotación de Información sirven como base para las distintas entidades de salud del País para poder dictaminar de manera más eficiente y segura si un paciente posee o no enfermedad hepática, teniendo como base el modelo desarrollado y las reglas de comportamiento. Teniendo en cuenta los diferentes valores de los distintos atributos identificados y usados por este Proyecto, se puede determinar el diagnostico del Paciente.

Hay atributos importantes a evaluar teniendo en cuenta el modelo desarrollado, por ejemplo la relación albumina globulina (A/G), dado que si es menor a 0,2 ya se puede detectar la presencia de enfermedad. Otros factores importantes son el Aspartato aminotransferas y bilirrubina total, a los cuales hay que prestarles atención para posteriores análisis y evaluación de otros atributos.

Por medio de este modelo se pueden analizar los distintos valores tomados al paciente de los ejes dispuestos (proteínas, relación A/G, etc) y evaluar, por medio de las reglas desarrolladas, si el paciente posee o no alguna enfermedad. Este modelo es fácil de usar e interpretar.

Este Proyecto de Explotación de Información ha cumplido su principal meta de brindarle conocimiento a las empresas que se encarguen de la salud para que puedan tomar decisiones más acertadas y certeras, disminuyendo el riesgo y la incertidumbre y que los doctores o encargados de brindarle al paciente una respuesta, puedan diagnosticar a los mismos de forma más rápida y concreta. Técnicamente hablando, el Proyecto de Explotación de Información desarrollado en este

trabajo, cumplió la meta de lograr a partir de los datos obtenidos y limpios, saber bajo que ciertas reglas si un paciente que se atiende en alguna entidad de salud posee o no alguna enfermedad hepática.

La información brindada y descubierta es útil para todo Organismo dedicado a la salud que quiera diagnosticar de forma más rápida y eficiente a sus pacientes.

Tal como se menciona en secciones anteriores, el éxito de este proyecto estaría alcanzado si el Ministerio de Salud puede nutrir a sus organismos de información y conocimiento necesario para que puedan tomar decisiones de manera precisa y que muchos pacientes puedan ser diagnosticados de manera exacta y sin errores que pueden acarrear más problemas. Por este motivo se puede dictaminar que este Proyecto ha alcanzándolo el criterio de éxito descripto.

Dado que el criterio de éxito del negocio en sí es uno solo (brindarle a las distintas entidades y organizaciones de salud conocimiento y reglas de comportamiento para que puedan brindarle a sus pacientes un diagnostico más certero y seguro), no amerita la creación de un ranking de resultados con respecto a los mismos.

Las metas iniciales de la aplicación se han alcanzado en el presente Proyecto, ya que los resultados obtenidos ayudarán al diagnostico de enfermedades hepáticas.

Como objetivos de negocio que deban ser evaluados mas tarde en el Proyecto o en un nuevo Proyecto se define el análisis de pacientes con Hepatitis A, B y C, las cuales son las más comunes de este ramo de enfermedades, para dictaminar los atributos más importantes y sus valores a la hora de poseer y contraer dicha enfermedad.

Teniendo en cuenta las hipótesis y metas descriptas en secciones anteriores, se ha confirmado que pacientes con menor edad son menos propensos a contraer alguna enfermedad hepática, como así también que existe una relación muy directa entre la edad y el diagnostico final. Por otro lado, teniendo en cuenta la cantidad de bilirrubina detectada, se ha confirmado que una porción menor a 10 determina sin ninguna evaluación más que el paciente posee enfermedad hepática, no así si es mayor a ese valor.

Se han aprobado tanto el modelo diseñado como el alternativo en el presente Proyecto ya que se estableció que cumple con las metas propuestas en el mismo.

5.3.2.5.2. Revisar el proceso

Esta segunda actividad se realiza dentro del Modelo de Ciclo de Vida en la fase Revisión Proceso. A continuación se desarrolla para el Proyecto de E.I. actual.

Mediante el uso de la herramienta Tanagra, se ha realizado el Proyecto de Explotación de Información requerido. Se han configurado los parámetros, y por medio del proceso de Descubrimiento de Reglas de Comportamiento, se llevo a cabo la utilización del algoritmo de

inducción TDIDT C4.5 para presentar dichas reglas. Teniendo en cuenta la utilización del algoritmo C4.5 para llevar a cabo el descubrimiento de reglas de comportamiento, se puede dictaminar que su utilización ha sido llevada a cabo de forma optima, mostrando resultados claros y concisos.

- Bilirrubina_Directa < 1,2500
 - Alanina_Aminotransferasa < 66,5000
 - Fosfatasa_Alcaldina < 601,0000
 - Edad < 66,5000
 - Edad < 59,0000
 - Edad < 25,5000
 - Aspartato_Aminotransferasa < 41,5000 then Diagnostico_Final = **Negativo** (70,59 % of 34 examples)
 - Aspartato_Aminotransferasa >= 41,5000 then Diagnostico_Final = **Positivo** (60,00 % of 15 examples)
 - Edad >= 25,5000
 - Fosfatasa_Alcaldina < 236,5000
 - Bilirrubina_Total < 1,6500
 - Fosfatasa_Alcaldina < 132,5000 then Diagnostico_Final = **Positivo** (80,00 % of 15 examples)
 - Fosfatasa_Alcaldina >= 132,5000
 - Proteinas < 5,1500 then Diagnostico_Final = **Positivo** (68,75 % of 16 examples)
 - Proteinas >= 5,1500
 - Albúmina < 4,3500
 - Sexo < 1,5000
 - Bilirrubina_Directa < 0,1500 then Diagnostico_Final = **Positivo** (72,73 % of 11 examples)
 - Bilirrubina_Directa >= 0,1500
 - Edad < 45,5000
 - Aspartato_Aminotransferasa < 25,0000 then Diagnostico_Final = **Negativo** (100,00 % of 13 examples)
 - Aspartato_Aminotransferasa >= 25,0000
 - Fosfatasa_Alcaldina < 183,0000 then Diagnostico_Final = **Negativo** (80,00 % of 15 examples)
 - Fosfatasa_Alcaldina >= 183,0000 then Diagnostico_Final = **Positivo** (72,73 % of 11 examples)
 - Edad >= 45,5000 then Diagnostico_Final = **Positivo** (62,50 % of 24 examples)
 - Sexo >= 1,5000
 - Edad < 46,5000 then Diagnostico_Final = **Positivo** (78,57 % of 28 examples)
 - Edad >= 46,5000 then Diagnostico_Final = **Negativo** (61,11 % of 18 examples)
 - Albúmina >= 4,3500 then Diagnostico_Final = **Negativo** (69,23 % of 13 examples)
 - Bilirrubina_Total >= 1,6500 then Diagnostico_Final = **Positivo** (80,95 % of 21 examples)
 - Fosfatasa_Alcaldina >= 236,5000
 - Edad < 45,5000
 - Proteinas < 6,6500 then Diagnostico_Final = **Negativo** (70,00 % of 10 examples)
 - Proteinas >= 6,6500 then Diagnostico_Final = **Positivo** (92,31 % of 13 examples)
 - Edad >= 45,5000 then Diagnostico_Final = **Positivo** (100,00 % of 18 examples)
 - Edad >= 59,0000 then Diagnostico_Final = **Negativo** (61,70 % of 47 examples)
 - Edad >= 66,5000 then Diagnostico_Final = **Positivo** (81,82 % of 33 examples)
 - Fosfatasa_Alcaldina >= 601,0000 then Diagnostico_Final = **Positivo** (100,00 % of 10 examples)
 - Alanina_Aminotransferasa >= 66,5000 then Diagnostico_Final = **Positivo** (89,39 % of 66 examples)
 - Bilirrubina_Directa >= 1,2500 then Diagnostico_Final = **Positivo** (95,27 % of 148 examples)

Fig. 5.10. Modelo con parámetros ajustados

En este proyecto actualmente desarrollado no se han detectado fallas, pasos faltantes, ni la necesidad de identificar posibles acciones alternativas.

Teniendo en cuenta lo desarrollado, el grupo de desarrollo de este Proyecto dictamina que los resultados obtenidos por el mismo han alcanzado los criterios del éxito y puede ser utilizado sin problemas por cualquier entidad de salud para el diagnóstico de enfermedades hepáticas.

5.3.2.5.3. Determinar próximos pasos

La última actividad de este SubProceso se realiza en la fase Determinación de Próximos Pasos y se desarrolla a continuación:

El resultado obtenido por medio de este Proyecto puede ser ampliado si se tomase menor “Min size of leaves”, aumentando así al detalle los posibles comportamiento de los valores para detectar la presencia o no de enfermedad.

Se recomienda realizar un periodo de prueba para evaluar los resultados del modelo desarrollado, para, en un momento de madurez, tomar como fuerte base las reglas construidas en el modelo, para así diagnosticar de manera mas rápida y eficiente al paciente.

No resulta necesario refinar el plan de proceso del actual Proyecto.

Se dictaminó que las futuras acciones a realizar por el equipo de desarrollo es llevar a cabo un seguimiento del Proyecto realizado para atender cualquier inquietud del cliente, asesorándolo y capacitándolo en lo que necesite. Se le planteó como alternativa al cliente, la posibilidad de mostrar más al detalle el modelo disminuyendo el “Min size of leaves” del algoritmo C4.5 por un periodo de tiempo, por si no lo conforma el nivel de abstracción desarrollado (mas allá de que ese nivel del primer modelo presentado haya quedado establecido en los requisitos)

5.3.2.6. Entrega

Este último SubProceso tiene una sola actividad dentro: Producir un Reporte Final.

5.3.2.6.1. Producir un reporte final

Esta última actividad se lleva a cabo en las últimas cuatro fases del Modelo de Ciclo de Vida en Espiral que posee el Proyecto y que se ha propuesto en este Trabajo: Desarrollo Plan, Plan Supervisión y Mantenimiento, Informe Producto y Revisión Proyecto. Finalizando el Proyecto se realiza la actividad mencionada:

Resulta necesario el desarrollo de un reporte o informe final el cual desarrolle los principales aspecto desarrollados en este trabajo, identificando objetivos, metas, interrogantes, hipotesis, el desarrollo del modelo, su análisis y evaluación.

Se han cumplido las metas pactadas y descritas en procesos anteriores sobre este Proyecto de Explotación de Información. Se brindará a partir de su funcionamiento herramientas y conocimiento a las empresas que se encarguen de la salud para que puedan tomar decisiones mas acertadas y

certeras, disminuyendo el riesgo y la incertidumbre y que los doctores o encargados de brindarle al paciente una respuesta, puedan diagnosticar a los mismos de forma más rápida y concreta. Técnicamente hablando, el Proyecto de Explotación de Información desarrollado en este trabajo logró saber y dar a conocer bajo ciertas reglas si un paciente que se atiende en alguna entidad de salud posee o no alguna enfermedad hepática. Esto produce mayor conocimiento a las organizaciones, mejorando la toma de decisiones.

Por otro lado, este proyecto cumplió las metas organizativas de la Consultora, dado que se cumplieron todos los pasos y proceso exitosamente y, al mismo tiempo, este Proyecto de Explotación de Información le brindará al Equipo de Desarrollo y Consultora experiencia para poder llevar a cabo en un futuro nuevos proyectos del mismo estilo, el cual servirá de base para analizar riesgos, acciones, inconvenientes, etc.

Los reportes estarán destinados en primer lugar al cliente (el Ministerio de Salud de la Republica Argentina) y en segundo lugar se guardara una copia en la Consultora, el cual queda archivado en el historial del Departamento Informático, que servirá como experiencia y documentación valiosa para nuevos proyectos que se presenten.

Como se ha dicho anteriormente, los grupos destinados a recibir la presentación final del proyecto es el Ministerio de Salud de la Republica Argentina que utilizara dicho proyecto para alimentarse de conocimiento y poder diagnosticar de manera más eficiente a sus pacientes; y en segundo lugar se almacenara el reporte final dentro de la consultora, para que quede como experiencia y se pueda aprovechar para futuros proyectos a contraer.

Para finalizar este capítulo cabe aclarar que mas allá de que se haya mencionado que los datos para aplicar la Prueba de Concepto fueron obtenidos de la región norteña de Argentina, los mismos han sido tomados realmente en base a casos de India. Se ha dicho que eso para darle realismo a la prueba.

En el próximo capítulo se llevan a cabo las Conclusiones obtenidas a partir del presente Trabajo Final de Licenciatura, indicando adicionalmente las futuras líneas de investigación a partir del mismo.

6. Conclusiones

En este capítulo se presentan las conclusiones obtenidas a partir del desarrollo de este Trabajo Final de Licenciatura, comenzando en primer lugar con los aportes realizados (sección 6.1.), para finalmente señalar las futuras líneas de investigación y desarrollo derivadas a partir del presente Trabajo (sección 6.2.).

6.1. Aportes del Trabajo Final de Licenciatura

En este Trabajo Final de Licenciatura se comprueba que si bien la metodología existente para Proyectos de Explotación de Información, CRISP-DM [Chapman *et al.*, 2000], es la más completa para desarrollo de este tipo de Proyectos, hay fases que no cubre tal como los Procesos de Administración. Por ello, se detecta necesaria la utilización del Modelo de Procesos desarrollado por Vanrell [2012], el cual le brinda un valor agregado a la metodología CRISP-DM [Chapman *et al.*, 2000] al incluirle los Procesos de Administración. Estos procesos de Administración tienen la función de recolectar información que le sea útil a los procesos de Desarrollo para aumentar la calidad de los mismos permitiendo realizar ajustes y mantener un estándar en la realización de distintos proyectos llevando a cabo la conversión de procesos caóticos y aislados en procesos repetitivos y controlados.

Además de esto, se detecta que la metodología CRISP-DM [Chapman *et al.*, 2000] no posee un Análisis enfocado a los Riesgos. Tal como ocurre en los Desarrollos Convencionales de Software, en los Proyectos de Explotación de Información también pueden ocurrir Riesgos y es fundamental detectarlos y tratarlos tempranamente para que no haya inconvenientes en el Desarrollo del Proyecto.

Finalmente, en este Trabajo Final de Licenciatura se desarrolla un Modelo de Ciclo de Vida que sirve como base para realizar Proyectos de Explotación de Información, tomando la mejora que ha hecho Vanrell [2012] a la metodología CRISP-DM [Chapman *et al.*, 2000] con los SubProcesos de Administración de Proyectos, y realizando un enfoque a los Riesgos tomando como base el Modelo de Ciclo de Vida en Espiral desarrollado por Boehm para Proyecto Convencionales [Boehm *et al.*, 1988].

El Modelo propuesto es superior a la Metodología CRISP-DM [Chapman *et al.*, 2000] ya que tiene en cuenta los SubProcesos de Administración de Proyectos que se realizan a lo largo de todo el desarrollo del mismo, y principalmente le agrega un Análisis enfocado a los Riesgos que dentro de los mismos pueden ocurrir.

Por otro lado se construye un Mapa de Actividades, el cual se encuentra constituido por las actividades pertenecientes a los SubProcesos de Administración y Desarrollo de Proyecto correspondiente al Modelo de Procesos realizado por Vanrell [2012], y por las fases que se han conformado en el Modelo de Ciclo de Vida en Espiral construido en el Trabajo Final de Licenciatura. A partir de allí, se definieron las actividades del Modelo de Procesos que se realizan en cada fase del Modelo de Ciclo de Vida, justificando su elección.

Por otro lado, las respuestas a los interrogantes planteados en el Capítulo 3 pueden ser resumidas en las siguientes conclusiones:

Pregunta 1: Se logra llevar a cabo la construcción de un Modelo de Ciclo de Vida que es útil para el desarrollo de distintos Proyectos de Explotación de Información y que posee un Análisis de Riesgo, cuya característica no existía en las metodologías existentes hasta el momento.

Pregunta 2: Se definen los Riesgos que pueden llegar a aparecer en los distintos Proyectos de Explotación de Información. En primer lugar en el Capítulo 3 se mencionan Riesgos en general, y, en el Capítulo 5, se desarrolla un Proyecto de Explotación de Información (útil para desarrollar la validación de las propuestas hechas en el presente Trabajo) donde detectan diferentes Riesgos.

Pregunta 3: Finalmente, se lleva a cabo la construcción de un Mapa de Actividades que posee como ejes en primer lugar las actividades del Modelo de Procesos desarrollado por Vanrell [2012], tanto de Administración como de Desarrollo, y en segundo lugar las fases pertenecientes al Modelo de Ciclo de Vida propuesto en el presente Trabajo.

Por medio de la Prueba de Concepto del Capítulo 5, se puede llevar a cabo la puesta en práctica tanto del Modelo de Ciclo de Vida como del Mapa de Actividades aplicando dichas propuestas a un Proyecto de Explotación de Información real.

6.2. Futuras líneas de Trabajo

Se realiza una validación temprana e inicial tanto del Modelo de Ciclo de Vida como del Mapa de Actividades propuesto, teniendo como dominio un Proyecto de Explotación de Información real.

Se plantea como futuras líneas de trabajos poder llevar a cabo el desarrollo de una cantidad más amplia y variada de Proyectos de Explotación de Información para continuar evaluando la viabilidad de las Propuestas desarrolladas en el presente Trabajo Final de Licenciatura.

Si bien el Mapa de Actividades contempla todas las actividades pertenecientes al Modelo de Procesos desarrollado por Vanrell [2012], resulta un punto de interés seguir analizándolo para evaluar las actividades aplicadas en cada fase del Modelo de Ciclo de Vida. Además, por medio de casos de prueba, se puede verificar la falta de alguna actividad dentro de alguna fase del Modelo de Ciclo de Vida construido.

7. REFERENCIAS

- Boehm, B.W., C. Abts, A.W. Brown, S. Chulani, B.K. Clark, E. Horowitz, R. Madachy, D. Reifer, B. Steece.(2000). *Software Cost Estimation with COCOMO II*, Prentice-Hall, Englewood Cliffs.
- Boehm, Barry W (1988). *A Spiral Model of Software Development and Enhancement*. TRW Defense Systems Group.
- Britos, P. (2008). *Procesos de Explotación de Información Basados en Sistemas Inteligentes*. Tesis Doctoral Universidad Nacional de la Plata. La Plata. Buenos Aires. Argentina.
- Britos, P., Dieste, O., García-Martínez, R. (2008a). *Requirements Elicitation in Data Mining for Business Intelligence Projects*. En *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.
- Britos, P., Grosser, H., Rodríguez, D., Garcia-Martinez, R. (2008b). *Detecting Unusual Changes of Users Consumption*. En *Artificial Intelligence and Practice II*, Max Bramer Ed. (Boston: Springer), IFIP Series, 276: 297-306.
- Britos, P., Jiménez Rey, E., García-Martínez, E. (2008c). *Work in Progress: Programming Misunderstandings Discovering Process Based On Intelligent Data Mining Tools*. Proceedings 38th ASEE/IEEE Frontiers in Education Conference.
- 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.
- Cogliati, M., Britos, P., García-Martínez, R. (2006) *Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT* In: Bramer M (ed) *Artificial Intelligence in Theory and Practice*, Boston, Springer, IFIP Series 217:305-314
- D’Atri, M., Rodriguez, D., Garcia-Martínez, R. 2009. *Improving Pipeline Risk Models by Using Data Mining Techniques*. 24th World Gas Conference Proceedings (CD). Paper 663.
- Felgaer, P., Britos, P., and García-Martínez, R. (2006) *Prediction in Health Domain Using Bayesian Network Optimization Based on Induction Learning Techniques*. *Int. J. of Mod. Ph. C* 17(3): 447-455

- Flores, D., García-Martínez, R. Fernández, E., Merlino, H., Rodríguez, D., Britos, P. (2009). *Detección de Patrones para la Prevención de Daños y/o Averías en la Industria Automotriz*. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1021-1030. ISBN 978-897-24068-4-1.
- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. (2011b). *Towards an Information Mining Engineering*. En Software Engineering: Methods, Modeling, and Teaching (ed. C. Zapata Jaramillo). Editorial de la Universidad de Medellín (en prensa).
- García-Martínez, R., Lelli, R., Merlino, H., Cornachia, L., Rodríguez, D., Pytel, P., Arboleya, H. (2011a). *Ingeniería de Proyectos de Explotación de Información para PYMES*. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Pág. 253-257.
- Grosser, H., Britos, P., García-Martínez, R. (2005) *Detecting Fraud in Mobile Telephony Using Neural Networks*. LNAI 3533:613-615.
- Kuna, H., García Martínez, R. Villatoro, F. 2010. *Pattern Discovery in University Students Desertion Based on Data Mining*. Advances and Applications in Statistical Sciences Journal, 2(2): 275-286. ISSN 0974-6811
- Langseth, J., Vivatrat, N. (2003). *Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound*. Intelligent Enterprise 5(18): 34-41.
- 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.
- Lockheed-Martin, (1998). *LLC PRICE Systems, PRICE S Reference Manual Version 3.0*, Lockheed-Martin.
- Negash, S., Gray, P. (2008). *Business Intelligence*. En Handbook on Decision Support Systems 2, ed. F. Burstein y C. Holsapple (Heidelberg, Springer), Pág. 175-193.
- 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., 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.
- Schiefer, J., Jeng, J., Kapoor, S., Chowdhary, P. (2004). *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings 2004 IEEE International Conference on E-Commerce Technology. Pág. 162-169.
- Stefanovic, N., Majstorovic, V., Stefanovic, D. (2006). *Supply Chain Business Intelligence Model*. Proceedings 13th International Conference on Life Cycle Engineering. Pág. 613-618.
- Vanrell, J., Bertone, R., García Martínez, R. *Modelo de Proceso de Operación para Proyectos de Explotacion de Informacion*. Anuales del XVI Congreso Argentino de Ciencias de la Computación. Pág. 674-682. ISBN 978-950-9474-49-9.
- 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.
- 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/vanrelltrabajofinaldeespecialidad.pdf>. Página web vigente al 24/07/10.

