

METODOLOGÍA DE SELECCIÓN DE HERRAMIENTAS DE EXPLOTACION DE DATOS

Britos, P.¹, Fernández, E.^{1,2}, Ochoa, M.^{1,2}, Merlino, H.^{1,2}, Diez, E.^{1,2} y García Martínez, R.¹

¹Centro de Ingeniería del Software e Ingeniería del Conocimiento. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.

²Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. Universidad de Buenos Aires

pbritos@itba.edu.ar

Resumen: En el proceso de explotación de datos existe el riesgo de detectar en forma tardía que la herramienta elegida sea inadecuada para cumplir con los objetivos de negocio establecido. Esto trae como consecuencia la pérdida de tiempo y dinero a la organización. En este trabajo se presenta una propuesta de metodología que permite seleccionar una herramienta de explotación de datos a partir de la caracterización de un grupo de ellas.

1. Introducción

La importancia del impacto de una herramienta de explotación de datos en la organización y la inversión que la misma debe hacer en términos económicos, hacen que el proceso de selección de la herramienta sea un tema crítico. Se debe tener en cuenta también que no es una tarea que se haga frecuentemente y que se espera un determinado retorno de la inversión. La metodología que se propone en este trabajo apunta a organizar el proceso de selección de una herramienta para que la organización pueda escoger la que mejor se adapte a sus requisitos basándose no solo en cuestiones económicas, sino también en las necesidades propias del negocio. A la fecha se ha trabajado en abordajes metodológicos para: selección de ciclos de vida [Rossi, 2001], selección de actividades para el mapa de actividades en un proyecto de software [Diez *et al*, 2003], selección de herramientas de desarrollo de sistemas expertos [García-Martínez y Britos, 2004], selección de herramientas ERP [Chiesa, 2004], pero esta ausente el abordaje de metodologías que contribuyan a la selección de herramientas de explotación de datos. La existencia de proveedores diversos y herramientas de uso libre que permiten realizar la tarea sin contar con un método objetivo de selección de herramientas puede conllevar la adquisición de una herramienta inadecuada, lo cual trae como consecuencias: (a) pérdida de tiempo y dinero a la organización y (b) el incremento del riesgo de no cumplir con los objetivos de negocio establecido.

2. Resolución del problema

A continuación se presenta una propuesta de metodología que permite seleccionar una herramienta de explotación de datos de forma objetiva. Dicha metodología esta compuesta por las siguientes fases: Fase 1: Documentar necesidad, Fase 2: Análisis de la necesidad, Fase 3: Búsqueda en el mercado, Fase 4: Contacto con los proveedores, Fase 5: Entrevista con posibles candidatos y recopilación de información, Fase 6: Armado de informe de criterios a tener en cuenta, Fase 7: Evaluar los candidatos, Fase 8: Demostración del producto y Fase 9: Evaluación final de los productos.

2.1. Fase 1: Documentar necesidad

En esta fase se define y establece el marco general de referencia para la selección de una herramienta de explotación de datos. Los aspectos que se deben considerar son:

- Definir áreas y funciones de la organización que se involucrarán con la herramienta de explotación de datos.
- Definir los objetivos que se pretenden lograr a través de la explotación de información.

2.2. Fase 2: Análisis de necesidad

El objetivo de esta fase es documentar las características del negocio que la herramienta de explotación de datos debe atender con relación a: áreas de la organización involucradas, procesos de negocios alcanzados y costo máximo que se pagará por la adquisición de la misma. El objetivo es identificar los requerimientos para la búsqueda de proveedores. Esta fase se orientará a describir una herramienta que se adapte lo mejor posible a los objetivos del negocio de la organización, es decir no pagar un precio muy elevado por una herramienta que se usará en un 10% de su potencial, ni por otro lado comprar una herramienta que no aporte los métodos necesarios o que resulte obsoleta al primer intento de ampliación.

2.3. Fase 3: Búsqueda en el mercado

El objetivo de esta fase es la búsqueda en el mercado de herramientas de explotación de datos disponibles, para lo cual se sugiere consultar en Internet, concurrir a exposiciones de software, adquirir revistas del rubro, identificar bibliografía sobre el tema y consultar con profesionales en otras organizaciones. Este proceso de búsqueda debe resumirse en un informe que detalle los proveedores encontrados.

2.4. Fase 4: Contacto con proveedores

El objetivo de esta fase es contactar a cada proveedor identificado en la fase anterior y solicitarle información de la herramienta. No es necesario todavía concertar entrevistas, el objetivo es recopilar la mayor cantidad de información de cada producto. Se sugiere reducir la cantidad de candidatos a 5, ya que se deberá llevar a cabo un estudio más profundo de cada uno. Esto incluye entre otras actividades: demostraciones de producto, visitas de los usuarios al proveedor, entrevistas con personal del proveedor y redacción de informes por cada producto.

2.5. Fase 5: Entrevistar posibles candidatos y recopilar información

En esta fase se conciertan entrevistas con cada proveedor con el propósito de completar la información faltante de cada producto y proveedor en orden a permitir una adecuada comparación sobre especificaciones técnicas del sistema, descripción de los módulos que lo componen, funcionalidad de cada módulo, catálogos, artículos o trabajos de experiencias de implementaciones en otras organizaciones. Para terminar esta fase se organiza la información verificando que los datos recopilados sean homogéneos para facilitar la comparación. Se preparará un informe por cada herramienta donde figure la presentación institucional de cada proveedor y un resumen de las características de las mismas.

2.6. Fase 6: Armado de informe de criterios a tener en cuenta

El objetivo de esta fase es desarrollar un informe con puntos de comparación ponderados que se adecue a las necesidades de la organización y que será la base de trabajo para las tareas posteriores y para la selección final. Para la comparación y selección de un producto es necesario tener un informe de criterios ponderados y puntos de comparación comunes, teniendo en cuenta que se han identificado diferentes aspectos que deben ser evaluados en el proceso de selección. Un informe con criterios de ponderación puede ser usado como modelo (ver plantillas en sección 2.9 y 2.10), el mismo debe ser adaptado a las necesidades particulares de la organización y los explotadores de información (dataminers), verificando que los aspectos seleccionados se puedan aplicar a la organización en cuestión y que la ponderación sugerida es adecuada a la misma. Los criterios del informe son agrupados en cuatro categorías o grupos, que también serán ponderadas, a continuación se detallan los grupos:

1. **Características técnico-funcionales de la herramienta:** bajo esta categoría se agrupan todos los criterios a evaluar que están ligados a las características técnicas y funcionales de la herramienta.
2. **Características del proveedor:** aquellos criterios de evaluación que hacen a la organización proveedora como: evolución y crecimiento, facturación anual, ubicación geográfica, otros clientes y experiencia. Es importante evaluar la solidez del proveedor, ya que si el proveedor deja de existir la organización se queda con un sistema sin mantenimiento ni posibilidad de evolución.
3. **Características del servicio:** en estos aspectos se evalúa puntos específicos del servicio que brinda el proveedor sobre la herramienta.
4. **Características económicas:** son aquellos relacionados con costos de licencias y de servicio de mantenimiento de la herramienta.

Para armar el informe de criterios se sugiere seguir los siguientes pasos:

1. Armar el informe de criterios que mejor aplique a la organización, tomando como modelo los criterios propuestos, con los conocimientos adquiridos de las herramientas en función de la información recopilada y el informe de las necesidades definidas.
2. Dividir los criterios en cuatro grupos dependiendo de las características: técnico-funcionales de la herramienta, del proveedor, del servicio, y económicas.
3. Ponderar cada criterio según su impacto dentro del grupo. La suma de las ponderaciones de cada grupo debe ser igual a 100, siendo la suma de todos los criterios igual a 400. (Ver cuestionario propuesto en sección 2.9).

2.7. Fase 7: Evaluar los candidatos

En esta fase el equipo debe concertar nuevas entrevistas con los proveedores en orden a formular solicitud de propuesta técnica y económica que completa el punto armado en el punto anterior. Se recomienda visitar las oficinas del proveedor, concertar reuniones con personal comercial y técnico para tener distintas visiones del producto. Para completar el informe cada criterio será calificado con un valor de 1 a 4, en la columna valor "Y" del cuestionario propuesto sección 3 (siendo 1= Malo, 2 = Regular, 3 = Bueno, 4 = Muy Bueno). Cada valor en la columna "Y" será multiplicado por el factor en la columna "Pond X" y se colocará en la columna "Pond X*Y". La sumatoria de la columna "Pond X*Y" se la multiplicará por la ponderación del grupo y dividirá por 100, así se

obtendrá la ponderación del grupo en general. Se repite esta operación para los 4 grupos en evaluación y para todas las herramientas (Ver ejemplo en el modelo). Una vez completo el informe con todos los datos recolectados se compara las ponderaciones resultantes entre las distintas herramientas. A los informe armados para cada proveedor, se debe agregar el informe evaluado, el informe de ventajas y desventajas y una copia de la propuesta. Luego de esto es conveniente organizar una reunión de trabajo con el equipo de proyecto para presentar las opciones, discutir la evaluación, comparar los valores obtenidos y seleccionar los candidatos. Al finalizar esta actividad se deberán seleccionar los productos de los cuales se pedirá una demostración a los proveedores.

2.8. Fase 8: Demostración del producto

En este punto los proveedores mostrarán el producto a los usuarios seleccionados y ellos completarán en cada visita los cuestionarios armados (ver plantilla propuesta para la evaluación de productos en la sección 2.10). Los usuarios calificará cada criterio indicando en la columna de ponderación ("P") un valor del 0 a 5 según se explica en la cabecera del modelo propuesto. Al finalizar las visitas se recopilan los cuestionarios, se suman los puntajes de cada proveedor otorgado por cada encuestado y se arma un promedio de puntos obtenidos por cada producto. Se agrega al informe armado en la actividad anterior los cuestionarios y puntaje total obtenido por cada herramienta. Al terminar este punto se tiene un informe con la evaluación completa por candidato que incluye la información institucional, la propuesta, el informe de criterios ponderados, las encuestas evaluadas producto de las demostraciones, el cuadro de ventajas y desventajas y todo comentario e información adicional que se tenga del proveedor y del producto que se haya recopilado en estas actividades.

2.9. Fase 9: Evaluación de los productos

En esta fase se comparan los resultados obtenidos (criterios ponderados + demostración de productos) y se tomará aquel producto que sume mayor puntaje.

2.10. Plantillas propuestas para ponderación de criterios

Encabezado propuesto

Nombre de la herramienta:	
Proveedor:	
Evaluación:	1 = Malo, 2 = Regular, 3 = Bueno, 4 = Muy bueno

Cuestionario propuesto

Criterios de selección	Descripción	Pond X	Valor Y	Pond X* Y
1. Características técnicos – funcionales				
Metodología / Ciclo de vida soportado	Metodologías / ciclos de vida que soporta la herramienta para la explotación de datos (CRISP-DM, SEMMA, etc.)	3		
Adaptabilidad y flexibilidad para la toma de datos	Desde Bases de Datos	Cantidad de formatos soportado para la toma de datos desde bases de datos diversas.	8	
	Desde fuentes externas (word, excel, etc.)	Cantidad de formatos soportado para la toma de datos.	8	
Facilidad para integrar diferentes técnicas	Posibilidad de integrar diversas técnicas de explotación de datos	5		
Multi-lenguaje	Permite trabajar en distintos idiomas (tomando el idioma ingles como idioma principal).	2		
Técnicas usadas	Cantidad de técnicas que permiten la explotación de datos para el <u>logro de los objetivos del negocio</u> (redes neuronales, redes bayesianas, algoritmos de inducción, etc.)	18		
Herramientas de visualización y informe	Permite visualizar la salida de las distintas técnicas utilizadas en la explotación, así como la generación de informes.	12		

Criterios de selección	Descripción	Pond X	Valor Y	Pond X* Y
1. Características técnicas – funcionales				
Multiplataforma	No necesita una plataforma determinada, es posible que se ejecute en varias plataformas	5		
Instalación remota	Permite instalación y trabajo del personal técnico en forma remota, sin estar en el lugar físico en donde esta el servidor	5		
Múltiples usuarios	Trabaja con una estructura que permitan múltiples usuarios.	2		
Seguridad	Perfiles de seguridad por tipo de usuario	2		
Resguardo de información	Metodología para copias de resguardo y recuperación de datos	2		
Desarrollo de la interfaz	Tipo de interfaz de la herramienta	10		
Gestor de configuraciones	Posee herramientas que administran las distintas versiones generadas	8		
Documentación técnica	Documentos sobre el funcionamiento de la herramienta.	5		
Conectividad externa	Soporta conexiones externas del tipo: Internet, EDI, Accesos remotos	2		
Compatibilidad con correo electrónico	Permite derivar informes o gráficos por e-mail.	3		
TOTAL		100%	Z= \sum	
	Ponderación del grupo	40%	P1 = Z * 0,40	

2.- Características del proveedor				
Características del proveedor	Solidez del proveedor: evolución histórica, clientes, ganancias, cantidad de empleados.	30		
Perspectivas de evolución	Perspectivas del proveedor en el mercado deben ser buenas ya que si al proveedor le va mal compraremos una herramienta que quedará sin soporte.	10		
Ubicación	Ubicación de las oficinas. Soporte en la misma ciudad donde se ubican las oficinas.	30		
Otras implementaciones	Desde el punto de vista del proveedor: Otras organizaciones que usen la herramienta.	5		
	Desde el punto de vista de los clientes que utilizan la herramienta (se sugiere consulta presencial o telefónica)	5		
Confianza	Criterio no cuantificable que queda a criterio de los miembros del equipo.	20		
TOTAL		100%	Z = \sum	
	Ponderación del grupo	25%	P2 = Z * 0,25	

3.- Características del servicio				
Garantía del producto	Problemas que estarían cubiertos por el proveedor y casos de los cuales el proveedor no se haría responsable. Alcance de la garantía en tiempo, en aspectos funcionales y técnicos	30		
Upgrade	Averiguar cada cuanto tiempo saca una nueva versión al mercado. Tener en cuenta si uno debe migrar obligatoriamente a la nueva versión al salir al mercado. De no ser así consultar cuanto tiempo el proveedor soporta las versiones más antiguas.	20		
Licencia	Alcance de la licencia. Incluye el soporte post venta. Alcance del soporte.	30		
Soporte	Existe un helpdesk para problemas no reportados en el repositorio con un tiempo de respuesta aceptable y atención 24 hs.	20		
TOTAL		100%	Z = \sum	
	Ponderación del grupo	20%	P3 = Z * 0,20	

4.- Características económicas				
Costos de la herramienta	En función del presupuesto que se tiene y de los otros presupuestos recibidos evaluar del costo del sistema.	30		
Costo del hardware	En función de los requerimientos de hardware y de lo que ya posee la organización, evaluar el costo que implica adquirir el equipamiento necesario.	20		
Costo de software	En función de los requerimientos de software y de los que ya posee la organización, evaluar el costo que implica adquirir mas software.	20		
Licencias	¿Como se pagan las licencias, por única vez al momento de la compra; o cuando ya se implementó o una vez por año?	10		
Financiación	Existen políticas de financiación para la adquisición del producto	10		
Upgrade	Costo del Upgrade. ¿Se deben abonar nuevas licencias?	10		
TOTAL		100%	Z = \sum	
	Ponderación del grupo	- 15%	P4= Z * (- 0,15)	

El factor económico se toma como valor negativo, ya que de esta manera en la ponderación final pesa aquellos productos económicos con respecto a los más costosos.

Ponderaciones por aspecto		
P1: Técnicos y funcionales	40%	
P2: Del proveedor	25%	
P3: Del servicio	20%	
P4: Económico	- 15%	
TOTAL	100%	

Ventajas y Desventajas	
Ventajas	Reservar una sección del cuadro para ventajas generales que puedan surgir de entrevistas con organizaciones que ya usan la herramienta.
Desventajas	Reservar una sección del cuadro para desventajas generales que puedan surgir de entrevistas con organizaciones que ya usan la herramienta.

2.11. Plantillas propuestas para la evaluación del producto

Encabezado propuesto

Nombre del dataminer:
Fecha:
Proveedor:
Ponderación: 0 = Ítem no evaluado 1 = Ítem evaluado no soportado. 2 = Ítem evaluado soportado de manera incompleta 3 = Ítem evaluado soportado con necesidad de varias modificaciones factibles 4 = Ítem evaluado soportado de manera correcta 5 = Ítem evaluado soportado y provee de valor agregado al trabajo

Cuestionario propuesto

CRITERIOS	P
Compañía con múltiples filiales	
Multiplataforma simultanea	
Multilinguaje - varios idiomas	
Ayudas en pantalla en el idioma de trabajo de la organización	
Manuales en el idioma de trabajo de la organización	
Importación de datos de distintas fuentes	
Cantidad de técnicas de explotación utilizadas para lograr los objetivos del negocio	
Integración entre técnicas	
Metodologías / Ciclos de vida utilizados	
Sistema de informes y visualización de resultados	
Niveles de seguridad	
Apreciación global del producto	
Confianza en el proveedor y la herramienta	
Conocimiento del producto por parte del proveedor	
Calidad de atención	
Respuesta a las consultas	
Presentación general	
TOTAL	

3. Estudio de Caso

3.1. Fase 1: Documentar necesidad

La organización en cuestión es una estación meteorológica que pretende conocer cual va a ser la velocidad de los vientos en una zona determinada y cuales son las causas que originan dicha velocidad (esto se utilizará como fuente de información para establecer mecanismos alternativos para la producción de energía), y en función de esto poder así establecer la cantidad de energía eólica que puede producir la planta eléctrica de la zona.

3.2. Fase 2: Análisis de necesidad

La herramienta de explotación de datos debe:

- Predecir la velocidad del viento, para lo que se sugiere el uso de redes neuronales de retropropagación,
- Indicar las reglas de comportamiento que originan la velocidad del viento, para lo que se sugiere algoritmos de inducción, por ejemplo C5, y
- Predecir la cantidad de energía eólica que se producirá, para lo que se sugiere el uso de redes neuronales de retropropagación.

3.3. Fase 3: Búsqueda en el mercado

Se toman como candidatos a evaluar las siguientes herramientas:

- Clementine, cuyo proveedor es SPSS.
- MatLab, cuyo proveedor es Mathworks.
- Weka, cuya distribución es libre, y pertenece al Department of Computer Science, University of Waikato, New Zealand.

3.4. Fase 4: Contacto con proveedores

Se recabo información de las herramientas por los diversos sitios webs y el envío de información por correo electrónico de los proveedores en cuestión:

- Clementine, cuyo proveedor es SPSS, www.spss.com.ar.
- MatLab, cuyo proveedor es Mathworks, www.mathworks.com.
- Weka, cuya distribución es libre, <http://www.cs.waikato.ac.nz/~ml/weka/index.html>.

3.5. Fase 5: Entrevistar posibles candidatos y recopilar información

De acuerdo a las herramientas analizadas solo se pudo obtener una entrevista con el proveedor de Clementine (SPSS), ya que el único que cuenta con representación en la zona de residencia de la organización. Para el resto de las herramientas se tomo información de el website y/o se envió correos electrónicos solicitando mayor información de las mismas.

3.6. Fase 6: Armado de listado de criterios a tener en cuenta

Los criterios a tener en cuenta son los propuestos en los modelos descriptos.

3.7. Fase 7: Evaluar los candidatos

Criterios de selección		Descripción	Pond X	Clementine		MatLab		Weka	
				Valor Y	Pond X*Y	Valor Y	Pond X*Y	Valor Y	Pond X*Y
1. Aspectos técnicos – funcionales									
Metodología / Ciclo de vida soportado		Metodologías / ciclos de vida que soporta la herramienta para la explotación de datos (CRISP-DM, SEMMA, etc.)	3	4	12	--	--	--	--
Adaptabilidad y flexibilidad para la toma de datos	Desde Bases de Datos	Cantidad de formatos soportado para la toma de datos desde bases de datos diversas.	8	5	40	4	32	3	24
	Desde fuentes externas (Word, Excel, etc.)	Cantidad de formatos soportado para la toma de datos.	8	5	40	4	32	3	24
Facilidad para integrar diferentes técnicas		Posibilidad de integrar diversas técnicas de explotación de datos	5	3	15	3	15	1	5
Multi-lenguaje		Permite trabajar en distintos idiomas (tomando el idioma inglés como idioma principal).	2	--	--	--	--	--	--
Técnicas usadas		Cantidad de técnicas que permiten la explotación de datos para el <u>logro de los objetivos del negocio</u> (redes neuronales, redes bayesianas, algoritmos de inducción, etc.)	18	5	90	5	90	5	90
Herramientas de visualización y informe		Permite visualizar la salida de las distintas técnicas utilizadas en la explotación, así como la generación de informes.	12	5	60	5	60	3	36
Multiplataforma		No necesita una plataforma determinada, es posible que se ejecute en varias plataformas	5	--	--	2	10	2	10
Instalación remota		Permite instalación y trabajo del personal técnico en forma remota, sin estar en el lugar físico en donde esta el servidor	5	3	15	NE	--	--	--
Múltiples usuarios		Trabaja con una estructura que permitan múltiples usuarios.	2	5	10	NE	--	--	--
Seguridad		Perfiles de seguridad por tipo de usuario	2	2	4	NE	--	--	--
Resguardo de información		Metodología de para copias de resguardo y recuperación de datos	2	--	--	NE	--	--	--
Desarrollo de la interfaz		Tipo de interfaz de la herramienta	10	5	50	5	60	3	36
Gestor de configuraciones		Posee herramientas que administran las distintas versiones generadas	8	--	--	--	--	--	--
Documentación técnica		Documentos sobre el funcionamiento de la herramienta.	5	5	25	5	25	4	20
Conectividad externa		Soporta conexiones externas del tipo: Internet, EDI, Accesos remotos	2	5	10	NE	--	--	--
Compatibilidad con correo electrónico		Permite derivar informes o gráficos por e-mail.	3	--	--	NE	--	--	--
TOTAL			100%		371		324		245
Ponderación del grupo			40%	P1	148,4		129,6		98

Criterios de selección		Descripción	Pond X	Clementine		MatLab		Weka	
				Valor Y	Pond X*Y	Valor Y	Pond X*Y	Valor Y	Pond X*Y
2.- Aspectos sobre el proveedor									
Características del proveedor		Solidez del proveedor: evolución histórica, clientes, ganancias, cantidad de empleados.	30	5	150	5	150	--	--
Perspectivas de evolución		Perspectivas del proveedor en el mercado deben ser buenas ya que si al proveedor le va mal compraremos una herramienta que quedará sin soporte.	10	5	50	5	50	3	30
Ubicación		Ubicación de las oficinas. Soporte en la misma ciudad donde se ubican las oficinas.	30	5	150	1	30	1	30
Otras implementaciones		Desde el punto de vista del proveedor: Otras organizaciones que usen la herramienta.	5	1	5	1	5	--	--
		Desde el punto de vista de los clientes que utilizan la herramienta (se sugiere consulta presencial o telefónica)	5	1	5	1	5	--	--
Confianza		Criterio no cuantificable que queda a criterio de los miembros del equipo.	20	5	100	5	100	3	60

Criterios de selección	Descripción	Pond X	Clementine		MatLab		Weka	
			Valor Y	Pond X*Y	Valor Y	Pond X*Y	Valor Y	Pond X*Y
2.- Aspectos sobre el proveedor								
TOTAL		100%		470		350		120
	Ponderación del grupo	25%	P2	117,5		87,5		30

Criterios de selección	Descripción	Pond X	Clementine		MatLab		Weka	
			Valor Y	Pond X*Y	Valor Y	Pond X*Y	Valor Y	Pond X*Y
3.- Aspectos sobre el servicio								
Garantía del producto	Problemas que estarían cubiertos por el proveedor y casos de los cuales el proveedor no se haría responsable. Alcance de la garantía en tiempo, en aspectos funcionales y técnicos	30	4	120	NE	--	1	30
Upgrade	Averiguar cada cuanto tiempo saca una nueva versión al mercado. Tener en cuenta si uno debe migrar obligatoriamente a la nueva versión al salir al mercado. De no ser así consultar cuanto tiempo el proveedor soporta las versiones más antiguas.	20	3	60	NE	--	2	40
Licencia	Alcance de la licencia. Incluye el soporte post venta. Alcance del soporte.	30	3	90	NE	--	--	--
Soporte	Existe un helpdesk para problemas no reportados en el repositorio con un tiempo de respuesta aceptable y atención 24 hs.	20	4	80	NE	--	--	--
TOTAL		100%		350		NC		70
	Ponderación del grupo	20%	P3	70		NC		14

Criterios de selección	Descripción	Pond X	Clementine		MatLab		Weka	
			Valor Y	Pond X*Y	Valor Y	Pond X*Y	Valor Y	Pond X*Y
4.- Aspectos económicos								
Costos de la herramienta	En función del presupuesto que se tiene y de los otros presupuestos recibidos evaluar de el costo del sistema.	30	1	30	NE	--	--	0
Costo del hardware	En función de los requerimientos de hardware y de lo que ya posee la organización, evaluar el costo que implica adquirir el equipamiento necesario.	20	--	0	5	100	--	0
Costo de software	En función de los requerimientos de software y de los que ya posee la organización, evaluar el costo que implica adquirir más software.	20	--	0	5	100	--	0
Licencias	¿Como se pagan las licencias, por única vez al momento de la compra; o cuando ya se implementó o una vez por año?	10	2	20	NE	--	--	0
Financiación	Existen políticas de financiación para la adquisición del producto	10	3	30	NE	--	--	0
Upgrade	Costo del Upgrade. ¿Se deben abonar nuevas licencias?	10	3	30	NE	--	--	0
TOTAL		100%		110		200		0
	Ponderación del grupo	- 15%	P4	- 16,5		- 30		0

Nota: La sigla NE significa "No evaluado".

Ponderaciones por aspecto				
		Clementine	MatLab	Weka
P1: Técnicos y funcionales	40%	148,4	129,6	98,0
P2: Del proveedor	25%	117,5	87,5	30,0
P3: Del servicio	20%	70,0	NE	14,0
P4: Económico	- 15%	- 16,5	- 30,0	0,0
TOTAL	100%	319,4	187,1	142

Ventajas y Desventajas			
	Clementine	MatLab	Weka
Ventajas	Representante local. Excelente interfaz.	Excelente interfaz	Uso libre.
Desventajas	Costo	Carencia de representante zonal. Carencia de información vía Web.	Poco soporte.

3.8. Fase 8: Demostración de los productos

Nombre del dataminer: Paola Britos			
Fecha: Abril de 2005			
Ponderación:			
0 = Ítem no evaluado			
1 = Ítem evaluado no soportado			
2 = Ítem evaluado soportado de manera incompleta			
3 = Ítem evaluado soportado con necesidad de varias modificaciones factibles			
4 = Ítem evaluado soportado de manera correcta			
5 = Ítem evaluado soportado y provee de valor agregado al trabajo			
CRITERIOS	Clementine	MatLab	Weka
Multicompañía	4	4	1
Multiplataforma simultanea	1	1	1
Multilenguaje - varios idiomas	1	1	1
Ayudas en pantalla en el idioma de trabajo de la organización	1	1	1
Manuales en el idioma de trabajo de la organización	1	1	1
Importación de datos de distintas fuentes	4	4	2
Cantidad de técnicas de explotación utilizadas para lograr los objetivos del negocio	5	5	4
Integración entre técnicas	2	2	2
Metodologías / Ciclos de vida utilizados	2	1	1
Sistema de reportes y visualización de resultados	2	5	3
Niveles de seguridad	5	5	1
Apreciación global del producto	5	4	3
Confianza	4	4	3
Conocimiento del producto por parte del proveedor	5	5	4
Calidad de atención	5	0	0
Respuesta a las consultas	5	0	0
Presentación general	5	0	0
TOTAL	57	43	29

3.9. Fase 9: Evaluación de los productos

A continuación se evaluó los productos:

Criterios de ponderación:

Ponderaciones por aspecto		Clementine	MatLab	Weka
P1: Técnicos y funcionales	40%	148,4	129,6	98,0
P2: Del proveedor	25%	117,5	87,5	30,0
P3: Del servicio	20%	70,0	NE	14,0
P4: Económico	- 15%	- 16,5	- 30,0	0,0
TOTAL	100%	319,4	187,1	142

Demostración de la herramienta:

Clementine	MatLab	Weka
57	43	29

Total:

Criterio	Clementine	MatLab	Weka
Criterios	319,4	187,1	142,0
Demostración	57,0	43,0	29,0
TOTAL	376,4	230,1	171

4. Conclusiones

De acuerdo a los ítems anteriormente desarrollado se puede establecer, que la metodología planteada:

- Establece un marco para la evaluación de herramientas de explotación de datos.
- Establece métricas para la evaluación de herramientas de explotación de información dentro de la organización.
- Es flexible y adaptable a cualquier tipo de organización.
- Permite evaluar de forma ordenada y ecuánime a un conjunto de herramientas de explotación de información, sin perder el objetivo de negocio de la organización.

5. Bibliografía

- Chiesa, F. 2004. *Metodología para Selección de Sistemas ERP*. Reportes Técnicos en Ingeniería del Software. (6)1:17-37.
- Diez, E., Britos, P., Rossi, B. y García-Martínez, R. 2003. *Generación Asistida del Mapa de Actividades de Proyectos de Desarrollo de Software*. Reportes Técnicos en Ingeniería del Software. (5)1:13-18.
- García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. ISBN 987-1104-15-4. Editorial Nueva Librería. Buenos Aires.
- Métrica, 1990. *Metodología de planificación y desarrollo de sistemas de información MÉTRICA Versión 2*.1990.
- Rossi, B. 2001. *Sistema Experto de Ayuda para la Selección del Modelo de Ciclo de Vida* Tesis de Master en Ingeniería del Conocimiento. Facultad de Informática. Universidad Politécnica de Madrid. <http://www.itba.edu.ar/capis/webcapis/tesisdemagisterterminadas.htm>. Pagina web vigente al 03/07/05.