



**Universidad  
Alberto Hurtado**

**Facultad de Economía y Negocios**

**Departamento de Gestión y Negocios**

**DESARROLLO DE PROPUESTA DE IMPLEMENTACIÓN DE  
BIBLIOMINING EN BIBLIOTECAS ESCOLARES CRA**

**Seminario de grado para optar al grado de Licenciado en Gestión de  
Información, Bibliotecología y Archivística**

**Por**

**FRANCISCA ABUMOHOR MATAMALA**

**BASTIÁN BESOAIN SEPÚLVEDA**

**PABLO VILLARROEL MALLEA**

**Profesor guía: Claudio Escobar Arriagada**

**Santiago, Chile**

**2018**

## **DEDICATORIAS**

Dedico esta tesis a mis padres, Mabel Matamala Escobar y Roberto Abumohor Cáceres, por haberme formado, apoyado, educado y guiado para convertirme en la persona que soy hoy.

A Patricia Abumohor Cáceres la persona responsable de mi decisión de estudiar bibliotecología.

A mis primas Cintia y Pamela Lucero Matamala, en especial a Pamela, quien me entregó muchas de las herramientas que me permitieron llevar a cabo mis estudios escolares y universitarios de mejor manera.

A mis hermanos, Rodrigo y Claudio, por su constante apoyo y enseñanzas de vida.

*Francisca Paz Abumohor Matamala*

Dedico esta tesis a Rosa Alvarado, por su esfuerzo para mantener las bibliotecas CRA accesibles y activas, y por enseñarme a ser riguroso en el trabajo cuando te apasiona.

A Felipe B. por estar ahí siempre que lo necesitara.

A Hector L. y a Patricia Ortiz por impulsarme a exigirme más.

*Bastián Esteban Besoain Sepúlveda*

Dedico esta tesis a Beatriz Mallea y Hugo Villarroel, mis papás, por haberme apoyado en todo momento, por su paciencia y por permitirme estudiar esta carrera.

*Pablo Ignacio Villarroel Mallea*

## **AGRADECIMIENTOS**

Queremos comenzar agradeciendo a Claudio Enrique Escobar Arriagada, nuestro profesor guía, por su incansable paciencia, dedicación y amor a su trabajo. Por acompañarnos durante 3 años en la carrera y en todo el proceso de creación de este seminario. Le agradecemos su disposición, sus ganas de trabajar y su alegría que nos animaron para no decaer y ayudarnos a levantarnos cuando faltaban fuerzas. Sin él nada de esto sería posible.

A la universidad, por darle un espacio a la Carrera Gestión de Información, Bibliotecología y Archivística.

## Tabla de contenidos

Resumen	5
Abstract	6
Capítulo 1: Presentación	10
1.1 Problema de investigación	10
1.2 Justificación	11
1.3 Objetivos	12
1.4 Metodología	12
Capítulo 2: Marco Teórico	14
1. Centro de Recursos para el Aprendizaje	14
1.1 Orígenes de las bibliotecas escolares CRA	14
1.2 Servicios	15
1.3 Desafíos	16
2. Minería de datos	20
2.1 Historia	22
2.2.2 CRISP-DM	25
2.2.3 CMIN	27
2.2.4 SEMMA	29
2.3.1.2 SAS	34
2.3.1.1 R	35
2.3.2.2 Python	37
2.4 Usos	38
3. Bibliomining	43
3.2 Usos	44
Gestión económica: selección y adquisición de fondos	46

3.1 Metodología	48
3.2 Tecnologías asociadas	49
3.2.1 Programas	49
3.2.2 Lenguajes	49
3.3 Propuesta de implementación	49
Evaluación de los resultados	55
Conclusiones	61
Sugerencias	61
Bibliografía	63
Glosario	66

## **Resumen**

El presente documento es un informe que detalla una propuesta de implementación de bibliomining para las bibliotecas escolares CRA en Chile. Expone desde las principales dificultades cotidianas que se presentan en las bibliotecas de este tipo hasta la importancia de la aplicación de la minería de datos enfocada al ámbito bibliotecológico, poniendo énfasis sobre el porqué se convierte en necesidad la idea de la mencionada propuesta de implementación. Por supuesto, se mencionan también las ventajas que el uso de esta herramienta pueda conllevar, así como también los elementos necesarios para que su uso y elaboración sean optimizados, una vez entendida la problemática que se presenta en una organización que carece de este aporte. El trabajo se organiza de tal manera que todas las temáticas son abordadas desde puntos sumamente estructurales, intentando ayudar a los lectores a contextualizar y conceptualizar las nociones e ideas con las que no estén familiarizados. Para complementar esto último, el informe también cuenta con un glosario que está a servicio y disposición de los lectores.

La propuesta de implementación de bibliomining para las bibliotecas escolares CRA en Chile expuesta en este informe está concebida bajo el modelo de CRISP-DM, el cual opera desde hace al menos dos décadas en los campos de la minería de datos y sus procesos analíticos. De la misma forma, Rapid Miner es el programa seleccionado para la propuesta, basado en las características que favorecen su aplicación y uso en las organizaciones de esta índole.

### **Palabras clave:**

Bibliomining, Minería de datos, Rapid Miner, CRISP-DM, Bibliotecas escolares, CRA

## **Abstract**

The following document is report that discusses a bibliomining implementation's proposal for school libraries in Chile. Expose main daily difficulties that a school library may suffer. Talks about how important the application of data mining can be applied to the world of librarianship, showing how relevant is the request of the bibliomining implementation's proposal for school libraries. In this report the advantages of using this tool are mentioned, as well as the necessary elements that allow its use and elaboration to be optimized, once the problematic has been understood in an institution that doesn't have this input.

This paper it's organized in order to approach every topic since structural points, helping the audience to contextualize and conceptualize the ideas that might not be familiar. To complement that previously mentioned point, this report has a glossary that cans the useful.

The bibliomining implementation's proposal for school libraries in Chile is backed up in this report by the CRIPS-DM model, which operates at least two decades ago in the data mining field and its analytic process.

In the same way, Rapid Miner is the chosen program for the proposal. Rapid Miner has a few characteristics that bring on its application and use for institutions of this nature.

**Keywords:** Bibliomining, Data mining, Rapid Miner, CRISP-DM, School libraries

## Introducción

La era digital se ha caracterizado por el uso de herramientas automatizadas en distintos procesos de las bibliotecas, con el objetivo principal de aumentar la producción y mejorar el acceso a información relevante, lo que ha tenido como consecuencia un crecimiento vertiginoso de las bases de datos a nivel organizacional, reflejando información de sus usuarios, historial de circulación, recursos en la colección y patrones de búsqueda (Koenig, 1985)<sup>1</sup>.

Comprendiendo lo mencionado, no es de extrañar que exista un creciente número de organizaciones, principalmente empresariales, que han reconocido la importancia de explotar estas fuentes de datos, a través de métodos informáticos desarrollados para extraer información relevante y analizarla, con el fin último de mejorar los procesos de la organización (López, 2013)<sup>2</sup>. Sin embargo no todas las organizaciones están preparadas, a nivel tecnológico, para procesar la cantidad de datos que poseen. A estas fuentes de datos se les conoce como Big Data (López).

Esta forma de interpretar los datos con el fin de mejorar procesos institucionales, se llama Minería de datos o en inglés Data Mining. Según el autor Marcelo de la Puente, la Minería de datos incluye el conjunto de técnicas para extraer información relevante almacenada en grandes volúmenes de datos, la cual podría resultar útil para modernizar procesos, implicando un análisis de datos para descubrir patrones de información, nuevo conocimiento entre otros (2010)<sup>3</sup>.

Adaptada al ámbito de las bibliotecas, surge el concepto de “bibliomining”, para hacer referencia a la técnica de minería de datos que procesa la información generada y almacenada por éstas. Además, el bibliomining tiene una estrecha relación con la bibliometría y la documentación, en un sentido en que aporta a la medición y el análisis de datos mediante métodos estadísticos, en este caso, para ayudar a definir tendencias o patrones de uso de los recursos electrónicos en bibliotecas y unidades de información (Quiroz, 2012)<sup>4</sup>.

---

<sup>1</sup> Koenig, M. E. D. (1985). Bibliographic Information Retrieval Systems and Database Management Systems. *Information Technology and Libraries*, 4, 247-272.

<sup>2</sup> López García, D. (2013). *Análisis de las posibilidades de uso de Big Data en las organizaciones* (Master's thesis). Disponible en: <https://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADa.pdf?sequence=1>.

<sup>3</sup> De-la-Puente, M. (2010). Bibliominería: bibliometría y minería de datos.

<sup>4</sup> Quiroz Gil, N. L., & Valencia, C. A. (2012). Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim. *Revista Interamericana de Bibliotecología*, 35(1).

En cuanto a los posibles beneficios que puede significar la aplicación de “bibliomining”, se encuentran los siguientes:

**Optimización de servicios:** Ayuda a los administradores a reorganizar el contenido de la biblioteca, autoridades e interfaces (Papatheodorou, 2003)<sup>5</sup>. También, a través del análisis de uso del OPAC, se puede estudiar la manera en la que los usuarios llegan a los registros, siendo de especial utilidad para ver qué términos utilizan, comparándolos con el lenguaje documental usado por la biblioteca, a modo de enriquecer la descripción temática y acercarla a la realidad del usuario (Candás, 2006)<sup>6</sup>.

**Apoyo a la toma de decisiones en distintas áreas de desempeño en las bibliotecas:** Una vez procesados los datos, se pueden establecer indicadores que den cuenta de la demanda de recursos, claridad sobre la necesidad, preferencia y uso por parte de quienes los consultan. Entre algunas de las áreas que se pueden beneficiar de estos servicios se encuentran la distribución del presupuesto y el desarrollo de colecciones, minimizando el riesgo en las inversiones y la ejecución del presupuesto en recursos no adecuados o de bajo uso (Quiroz).

**Personalización de los sistemas digitales:** Ayuda a los usuarios a identificar información de su interés por recomendación de materias similares, así como también entregando información beneficiosa, determinada por el tipo de necesidad de información que posea el usuario. Lo que supondría una mejora de los tan útiles sistemas de DSI.

**Evaluación bibliotecaria:** Proporciona datos directos y objetivos de los usuarios y de la utilidad que le dan a la biblioteca, frente a otros métodos con mayor o menor grado de parcialidad como las encuestas, entrevistas, el método Delphi y el incidente crítico. Estos indicadores pueden ser relacionados entre sí para la obtención de nuevos indicadores, que al ser estudiados pueden reflejar patrones de comportamiento de la actividad (Candás).

---

<sup>5</sup> Papatheodorou, Christos; Kapidakis, Sarantos; Sfakakis, Michalis; Vassilou, Alexandra (2003). "Mining user communities in digital libraries". *Information technology and libraries*, December, vol. 22, no. 4, p. 152-157.

<sup>6</sup> Candás Romero, Jorge (2006). "Minería de datos en bibliotecas : bibliominería". *BiD: textos universitaris de biblioteconomia i documentació*, núm. 17 (deseembre) . <<http://bid.ub.edu/17canda2.htm>>. [Consulta: 25-04-2018].



## Capítulo 1: Presentación

### 1.1 Problema de investigación

Los Centros de Recursos para el Aprendizaje (CRA de ahora en adelante) en Chile se establecen como el espacio especializado dentro de una institución escolar, privada o pública, con la misión de fomentar el interés por la información, la lectura y el conocimiento a través de espacios creativos de encuentro y aprendizaje para la comunidad escolar, así como también la generación continua de conocimiento por parte de los usuarios (Bibliotecas CRA)<sup>7</sup>.

En el caso puntual de Chile, los recursos económicos, humanos y de infraestructura escasean, siendo enfocados principalmente a mantener equipos tecnológicos, desarrollar colecciones y a la contratación de personal, que no siempre son profesionales de la información. Esto sumado al avance de las tecnologías, con el surgimiento de buscadores sencillos, aplicaciones de lectura basadas en la nube y la variedad de enciclopedias online, han desplazado gran parte de las funciones tradicionales de la biblioteca (Ministerio de Educación, 2014)<sup>8</sup>.

En datos concretos obtenidos por la encuesta Censal de Bibliotecas<sup>9</sup> sobre las necesidades de apoyo que tienen las bibliotecas escolares, se observa que el 99,5% de las bibliotecas declara necesitar apoyo de algún tipo para cumplir con sus funciones. La demanda más generalizada es por la renovación y ampliación de la colección para estudiantes. Cualquiera sea el volumen de la colección, ésta se cataloga como insuficiente en cantidad o diversidad en todos los establecimientos con biblioteca.

De esta manera, se detectan dos necesidades principales en la biblioteca escolar CRA: la necesidad de invertir los recursos de manera inteligente y el deber de potenciar los procesos que tengan mayor impacto para el usuario y el desarrollo de su misión.

---

<sup>7</sup>Ministerio de Educación. Misión. Disponible en: <<http://www.bibliotecas-cra.cl/?q=mision>>.

<sup>8</sup>Ministerio de Educación, (2014). 20 años de bibliotecas escolares CRA. Disponible en: <[https://issuu.com/bibliotecas\\_cra/docs/20\\_anos\\_de\\_bibliotecas\\_escolares\\_cr\\_a14a10b868c559](https://issuu.com/bibliotecas_cra/docs/20_anos_de_bibliotecas_escolares_cr_a14a10b868c559)>.

<sup>9</sup> Ministerio de Educación. Aplicación de encuesta censal en las bibliotecas escolares cra de los establecimientos educacionales subvencionados del país. Disponible en: <[http://www.bibliotecas-cra.cl/sites/default/files/publicaciones/estudioencuesta\\_censal\\_2011.pdf](http://www.bibliotecas-cra.cl/sites/default/files/publicaciones/estudioencuesta_censal_2011.pdf)>

## 1.2 Justificación

La realidad de las bibliotecas escolares, si bien parece nebulosa, aplicando el uso de tecnologías de análisis de datos (bibliomining) es posible contravenir esta realidad. Ésto es debido a que se ha generado la política de utilización de sistemas de automatización de bibliotecas dentro de las bibliotecas escolares CRA, entre ellos el sistema Abies 2.0, “software desarrollado por el Ministerio de Educación, Cultura y Deportes de España que ha sido cedido y adaptado para las bibliotecas escolares CRA del Ministerio de Educación de Chile y se entrega gratuitamente a los CRA que la soliciten, permite ingresar recursos y usuarios, registrar préstamos y devoluciones, imprimir estadísticas” (Ministerio de Educación)<sup>10</sup>. Si bien su uso no es obligatorio, se ha convertido en el servicio oficial para el Ministerio de Educación y su distribución es gratuita.

Como ya fue mencionado, entre los diversos usos de bibliomining, se encuentra la evaluación bibliotecaria como método de recolección de datos, de esta forma se espera predecir el impacto de las actividades realizadas por la institución, de modo que sirva como método para identificar ventajas y desventajas para una futura toma de decisiones administrativas.

Para cumplir con estos objetivos, se propone el desarrollo de una metodología de extracción y tratamiento de datos específica para bibliotecas, así como la examinación de las tecnologías más apropiadas para su posterior implementación.

A partir de todo lo planteado anteriormente, se presenta el uso de minería de datos para optimizar recursos como respuesta a la problemática de investigación.

Entre los resultados esperados de la aplicación de una metodología bibliomining en bibliotecas escolares CRA, se encuentran: detectar necesidades, intereses y oportunidades, de modo que se invierta de forma más eficiente los recursos económicos.

---

<sup>10</sup>Ministerio de Educación. Software Abies 2.0. Disponible en: <<http://www.bibliotecas-cra.cl/abies2>>.

### 1.3 Objetivos

#### Objetivo general

- Desarrollar una propuesta de implementación para la aplicación de bibliomining en distintas áreas de una biblioteca escolar CRA, así como explorar el uso de tecnologías existentes y analizar información relevante para bibliotecas, con el fin de facilitar la toma de decisiones.

#### Objetivos específicos

- Identificar las fuentes de datos más relevantes generadas en bibliotecas escolares CRA
- Identificar y comparar metodologías de procesamiento de datos para la implementación de bibliomining.
- Discutir y determinar las ventajas y desventajas de la aplicación de *bibliomining* en una biblioteca escolar.
- Identificar y comparar tecnologías requeridas para la implementación de una metodología de procesamiento de datos en bibliotecas.

Como se observa, los objetivos son de plano teórico y carecerán de sustento práctico, esto producto de que la presente investigación sólo será exploratoria, dados factores como el tiempo con el que se cuenta para su desarrollo y la limitada información de un tema que se encuentra aún en florecimiento en el ámbito académico.

### 1.4 Metodología

Se realiza una investigación preliminar con el fin de conocer a grandes rasgos qué está sucediendo hoy en día en el marco de la minería de datos.

Posteriormente, se define el problema de investigación y, conforme a la lectura, su posible justificación y posterior solución. Ésto será respaldado con una bibliografía acorde al tema de investigación. Esta bibliografía será el resultado del tercer paso de la metodología, donde se recopila información relevante que servirá de apoyo en el estudio.

Algo muy importante es definir el alcance que tendrá la investigación. Su objetivo es otorgar claridad acerca de los márgenes investigativos y determinar el campo de aplicación.

Con posterioridad, se procede a establecer el objetivo general del estudio, seguido de los objetivos específicos. Con ellos comprendidos, se da paso a la creación de la metodología de investigación y el marco teórico. Éste último directamente relacionado con los objetivos específicos, el cual se desarrollará

a partir de la bibliografía que se recolectará durante toda la investigación. La búsqueda bibliográfica se realiza recopilando artículos de revistas, libros; en español e inglés entre los años 2015 y 2018. Sin embargo, se extenderá dicho parámetro en caso de no encontrar información pertinente y/o suficiente.

Finalmente, a partir de toda la investigación y el análisis llevado a cabo, se da paso a las conclusiones y posibles sugerencias que puedan ser de utilidad tanto para la carrera Gestión de Información, Bibliotecología y Archivística de la Universidad Alberto Hurtado, la comunidad bibliotecaria, como para futuras investigaciones que se enmarquen en el área de las bibliotecas CRA o la minería de datos.

## **Capítulo 2: Marco Teórico**

### **1. Centro de Recursos para el Aprendizaje**

Como es sabido, las bibliotecas componen un lugar importante dentro de los establecimientos educacionales, colaborando con material bibliográfico que ayudará y guiará a los estudiantes en la formación académica. A este material, la biblioteca escolar CRA ofrece además otro tipo de recursos, entre los que se pueden mencionar los juegos de mesa (los cuales ayudan a desarrollar habilidades sociales) y, en oportunidades, computadores que pueden aportar con la búsqueda de información y entretenimiento del alumnado. Y es por esto último, que “La Biblioteca Escolar CRA tiene como fin principal apoyar el proceso de enseñanza y aprendizaje (fomento a la lectura, apoyo curricular y educación de usuarios (...)) en un ambiente atractivo y acogedor”<sup>11</sup>, de manera que pueda contribuir al desarrollo de los estudiantes al mismo tiempo que implica un apoyo para el profesorado de la institución.

De esta manera, los Centros de Recursos para el Aprendizaje serán concebidos como una nueva manera de percibir las bibliotecas escolares, intentando captar la mayor cantidad de usuarios posibles, independiente del cargo que ocupen las personas en el establecimiento educacional. En pocas palabras, una biblioteca escolar CRA no es otra cosa que una modalidad de una biblioteca escolar, que promueve la participación de su público objetivo por medio de distintos medios de captación, y que buscan el desarrollo personal y académico del alumnado que forma parte de la comunidad escolar.

#### **1.1 Orígenes de las bibliotecas escolares CRA**

Los Centros de Recursos para el Aprendizaje son los resultados que vieron nacer las bibliotecas escolares en Chile que formaron parte del proyecto de Mejoramiento en la Calidad y Equidad de la Educación (MECE) en el año 1993, que se vio influenciado por otros movimientos de similares características a nivel mundial, que, en las últimas décadas del siglo XX, buscaban la transformación de una biblioteca estudiantil, fuera universitaria o escolar, en un Centro de Recursos para el Aprendizaje y la Investigación, basándose en la creciente ola tecnológica que acaecía.

El proyecto MECE contempló las implementaciones de diversos recursos que pueden ser audiovisuales, instrumentales o digitales entre otros, con el “propósito de convertir la biblioteca escolar CRA en un espacio de formación, información y recreación”<sup>12</sup>, con la intención extra de romper con los prejuicios

---

<sup>11</sup> MINEDUC, (2011). Estándares para las Bibliotecas Escolares CRA.

<sup>12</sup> MINEDUC, Bibliotecas Escolares CRA.

generados en las biblioteca, vistas como lugares de poca interacción, sin libre acceso y, en ocasiones, hasta de castigo.

Básicamente, las bibliotecas escolares previas a la incorporación de los CRA carecían de un programa, “un proyecto que, desde el Estado de Chile, fomentara, regulara y masificara su implementación (la de los CRA) en los establecimientos educativos.”<sup>13</sup> Esto último, debido a que durante un largo período de tiempo no se juzgó como necesaria la creación de una política pública relacionada con este tema, considerando además, que inclusive a las mismas bibliotecas les hacía falta un rol más protagónico. Entendiendo que hasta 1940 la tasa de alfabetismo en Chile era de un humilde 58,3%<sup>14</sup>, es comprensible la escasa vinculación del Estado con estos tópicos.

Es bajo esta premisa que las bibliotecas escolares CRA comienzan su incorporación al sistema educativo nacional recién en el año 1993, potenciado con la vertiente tecnológica que permitiera una captación relevante del público joven, adquiriendo una relevancia cada vez más significativa hasta convertirse en el modelo de las bibliotecas escolares actuales. No obstante, también es de relevancia aclarar que desde la creación de las bibliotecas escolares CRA hasta la fecha, no ha existido una clara documentación del desarrollo evolutivo que han contemplado a estas bibliotecas escolares, más allá de una mayor cobertura a nivel país y los medios tecnológicos con los que los CRA cuentan, lo que se traduce en un escaso seguimiento del tema.

## 1.2 Servicios

Como fue mencionado anteriormente, los Centros de Recursos para el Aprendizaje cuentan con servicios que aportan al desarrollo del estudiantado, contribuyendo a un óptimo proyecto educativo en pos del fomento a la lectura. Entre estos muchos servicios, el documento Estándares para las Bibliotecas Escolares CRA, elaborada en el año 2011 por el Ministerio de Educación, destaca los siguientes<sup>15</sup>:

- Servicio de préstamo y ayuda a los usuarios para leer, investigar y estudiar en el CRA, suponiendo un acceso libre a la colección por parte de los usuarios.
- Servicio al aula: préstamo de material y recursos para el proceso enseñanza-aprendizaje dentro de las salas de clase.

---

<sup>13</sup> MINEDUC, (2015). 20 Años de Bibliotecas CRA.

<sup>14</sup> Brunner, José Joaquín (2008). Problemas, realidades y desafíos de la Educación en Chile.

<sup>15</sup> MINEDUC, (2011). Estándares para las Bibliotecas Escolares CRA.

- Préstamo a domicilio: servicio que ofrece a los usuarios la posibilidad de pedir préstamo de material bibliográfico por un período estipulado de tiempo.
- Préstamo interbibliotecario: servicio cooperativo y colaborativo entre bibliotecas con el fin de complementar colecciones de ambas partes.
- Servicio de consulta o referencia: como su nombre indica, este servicio considera la orientación y guía a los usuarios en cuanto a la búsqueda de información, basándose en el perfil del usuario.
- Programas de gestión pedagógica: organización de actividades, educación de usuarios y difusión cultural, entre otras de las mecánicas para aportar con el desarrollo curricular del usuario.
- Servicio de entretenimiento adaptado a una comunidad escolar, esto implica desde juegos de mesa hasta puzzles y rompecabezas.
- Recursos audiovisuales

Como se aprecia, es amplia y variada la gama de servicios que ofrecen las bibliotecas escolares CRA, que va de la mano con el ya mencionado objetivo de captar al mayor número de usuarios posibles, intentando satisfacer las necesidades de información que conllevan, al tiempo que servicios como el préstamo a domicilio pueden influir aún más en la cantidad de usuarios para que formen parte de la comunidad, que no los obliga a estar físicamente en la biblioteca para hacer uso de su material bibliográfico.

### **1.3 Desafíos**

Lo primero a destacar es que, como se espera haber dejado claro, las bibliotecas escolares CRA están en desarrollo continuo, así como su programa está en plena consolidación. Además, no basta con el reconocimiento del programa y de la implementación, a lo largo del país, de las bibliotecas escolares CRA como un elemento básico de la educación chilena contemporánea, sino que urge una normativa a nivel nacional que dé forma legal a estas bibliotecas, tanto como programa como lugar de acceso a información y lectura en las escuelas nacionales.<sup>16</sup>

Se hace más que notoria también, la idea de profesionalizar la red de coordinadores y encargados CRA a lo largo de Chile. Para ello, se hace menester ofrecer un plan de formación permanente en el personal, promoviendo la especialización tanto en el área bibliotecológica como

---

<sup>16</sup> MINEDUC, (2015). 20 Años de Bibliotecas CRA.

pedagógica.<sup>17</sup> En otras palabras, lo que se busca es capacitar a encargados de las bibliotecas escolares CRA, de manera que se desenvuelvan correctamente en el área de la gestión de información, en caso de ser docentes, o bien, entenderse en materia pedagógica al tiempo que son bibliotecarios.

Otro de los desafíos a mencionar es la búsqueda de la igualdad de recursos en cuanto a la puesta en funcionamiento de las bibliotecas escolares CRA se trata, es decir, que la implementación de las bibliotecas escolares CRA aporten con herramientas a toda su comunidad educativa, al tiempo que no se convierta en un reflejo del nivel social, económico o cultural de la misma<sup>18</sup>. Este punto toca una fibra importante a la hora de hablar de cualquier política pública en el país, y es la notable brecha existente de los recursos económicos y culturales entre las personas que lo componen.

En cuarto lugar, se posiciona de manera importantísima la necesidad de fortalecimiento del uso de los recursos de aprendizaje, de la misma manera que elementos como el fomento a la lectura, la alfabetización informacional y una mayor orientación a la gestión curricular se convierten en los principales protagonistas de la época<sup>19</sup>. Entendiendo esto, el documento 20 años de Bibliotecas Escolares CRA del MINEDUC, ofrece puntos que supone mejorarán el programa de estos Centros de Recursos para el Aprendizaje<sup>20</sup>. Se destacan:

- Desarrollar una estructura regional y provincial: Encargado de la mejora y concretización de un sistema de seguimiento en que se monitoreen las funciones y el cumplimiento de los estándares de las bibliotecas escolares CRA a lo largo del país.
- Sistematización del trabajo del equipo encargado que evalúa y selecciona los recursos de aprendizaje que integrarán las bibliotecas escolares CRA.
- Contribuir, en el largo plazo, a una relación más enriquecedora entre las prácticas pedagógicas y prácticas de lectura, con el objetivo de dinamizar el proceso de aprendizaje por medio de la lectura.

---

<sup>17</sup> MINEDUC, (2015). 20 Años de Bibliotecas CRA.

<sup>18</sup> Pérez Kauscuhs, Gabriela (2015):“EL ROL DE LAS BIBLIOTECAS ESCOLARES CRA EN EL SISTEMA EDUCATIVO CHILENO: DESCRIPCIÓN, DESAFÍOS Y RECOMENDACIONES PARA SU MAYOR EFECTIVIDAD EN EL INCREMENTO DE OPORTUNIDADES EDUCATIVAS”

<sup>19</sup> MINEDUC, (2015). 20 Años de Bibliotecas CRA.

<sup>20</sup> MINEDUC, (2015). 20 Años de Bibliotecas CRA.

- Proveer a las bibliotecas escolares CRA con los recursos de aprendizaje del presente, de modo que ni el material bibliográfico ni las tecnologías y servicios digitales presenten bajas en su calidad. Optar por dispositivos de lectura y la correcta formación para el uso de éstos.
- Difundir y potenciar las publicaciones CRA a través de los medios de comunicación.
- Llegar de manera más efectiva a toda la comunidad educativa, poniendo un especial énfasis en el cuerpo directivo, integrando los recursos de la biblioteca escolar CRA en el proyecto educacional del establecimiento. También se entiende por “comunidad educativa” a padres y apoderados, que cumplen roles fundamentales en la formación de sus pupilos, colaborando con insertarlos en el hábito de la lectura.

Estos son los puntos que se necesitan afrontar con mayor énfasis en las bibliotecas escolares CRA, de modo que, es posible dilucidar, es un camino largo de recorrer. Si bien es cierto, el avance ha sido sustantivo, aún hay muchas cosas de prioridad por afinar.

#### **1.4 Desarrollo en Chile**

Como ya se ha hecho explícito, el desarrollo en Chile de las bibliotecas escolares CRA ofrece una amplia gama de servicios, incorporándose de manera efectiva en los establecimientos educacionales y logrando la captación de los estudiantes, haciéndose coprotagonistas del desarrollo educativo en Chile.

Como buena materia relacionada con la educación, las bibliotecas escolares CRA forman parte de los quehaceres del Ministerio de Educación, quien se encarga de darles forma a través de estándares y aplicando directrices que fiscalicen el proyecto de manera óptima. Es por esto que las bibliotecas escolares CRA cuentan con sitio web propio y sus estándares y normas están en línea, de modo que toda persona pueda acceder a documentos que le provean información respecto de estas crecientes bibliotecas escolares.

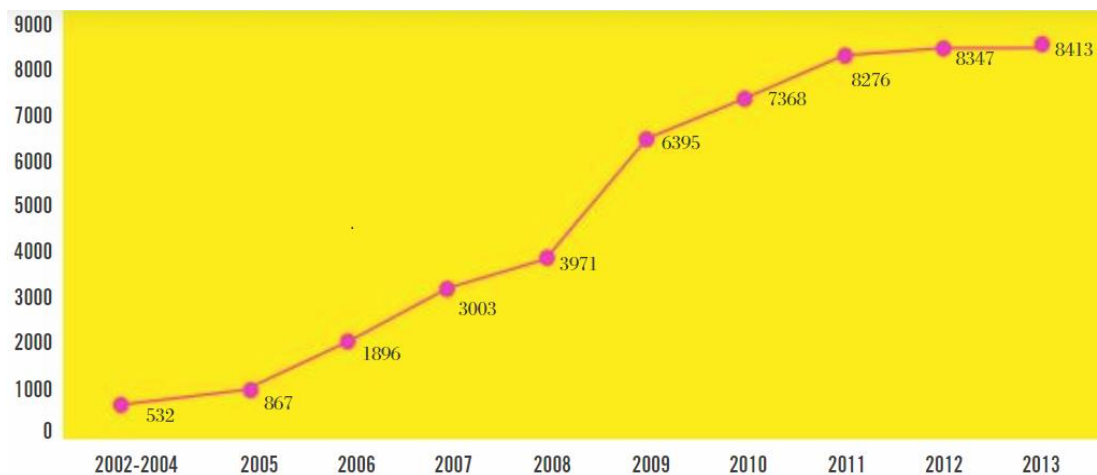


Fig. 1 Evolución Cronológica de la cobertura de las bibliotecas escolares CRA en la enseñanza básica. Fuente: Datos de Implementación Componente Bibliotecas Escolares CRA, UCE-MINEDUC).

Como la imagen indica, la implementación de las bibliotecas escolares CRA ha evidenciado un importante crecimiento a nivel país, que presenció su más importante alza hacia el año 2009, mientras que en los años posteriores se aprecia una evolución más pausada y menos cuantitativa, lo que se traduce en una sencilla pero firme estabilidad ascendente.

Actualmente, las bibliotecas escolares CRA se han hecho de tal relevancia en la educación nacional, que el Ministerio de Educación ofrece la posibilidad, a través de postulaciones, de que los establecimientos educacionales que carecen de una biblioteca escolar CRA, puedan recibir fondos destinados a la implementación de una. Las postulaciones son anuales y los documentos para llevarlos a cabo pueden ser solicitados en el sitio web institucional.

### 1.5 Principales fuentes de datos

Para la investigación, es lógico mencionar que es necesario contar con datos a estudiar. En el caso de las bibliotecas escolares CRA, es posible mencionar desde cuánto es el uso cotidiano de los estudiantes en tiempo, hasta, por ejemplo, el lugar geográfico en donde se encuentra ubicado el establecimiento educacional, o desde cuántos estudiantes frecuentan la biblioteca en relación al total de estudiantes del colegio; en qué horario la frecuentan, cuán equipada está la biblioteca escolar CRA objeto de estudio, así como su cantidad y calidad de recursos, el desempeño académico de los estudiantes que la frecuentan, cuánto utilizan el material bibliográfico en relación al audiovisual, si utilizan o no los servicios de préstamo a domicilio, etc.

Scott Nicholson, expone dos principales fuentes de datos que deberían ser consideradas: las fuentes de datos internas al sistema de la biblioteca y las

fuentes de datos externas al sistema de la biblioteca<sup>21</sup>. El primero, y como su nombre indica, focalizado en los datos que pueden extraerse desde la biblioteca, tales como los patrones observados en el sistema de biblioteca, y, a modo de ejemplo, pueden mencionarse el ordenamiento de las estanterías de la misma, o el funcionamiento de su sitio web, o el área en el cual la organización se desempeña, hacia dónde está orientada, su misión, su visión, entre otros. Las fuentes de datos externas, por su parte, son aquellas que pueden conseguirse a través de los factores externos a la biblioteca, y, como ejemplo, es posible destacar la localización geográfica de la misma y los perfiles de los usuarios registrados en la base de datos. Se destaca, en este punto, principalmente información demográfica.

## 2. Minería de datos

La minería de datos (De ahora en adelante nos referiremos como MD) agrupa todas las técnicas necesarias para la extracción de datos así como las técnicas para su posterior interpretación, con el objetivo de buscar patrones e información relevante y oculta dentro de los datos almacenados en grandes bases de datos (Sosa, Chesñevar y Sosa Bruchmann, 2016)<sup>22</sup>.

Normalmente se aplica MD en organizaciones que poseen grandes volúmenes de datos, preferentemente almacenada y estructurada en bases de datos digitales, para obtener información que permita la mejora de procesos de negocios, fungiendo de esta manera como herramienta de análisis y descubrimiento de conocimiento a partir del procesamiento de datos experimentales (De la Puente, 2010).

Un proceso normal de minería de datos implica una selección del conjunto de datos, un análisis de los mismos, una selección previa de las técnicas a aplicar, el proceso de extracción de conocimiento en sí mismo y la interpretación de los resultados obtenidos (De la Puente, 2010).

Martínez y Ocampo(2009)<sup>23</sup> identifican varias técnicas de clasificación de datos, algunas de ellas son los árboles de decisiones, clúster, redes

---

<sup>21</sup> Nicholson, Scott. (2004). The Bibliomining Process: Data Warehouse and Data Mining for Library Decision Making.

<sup>22</sup> Sosa, M. O., Chesñevar, C. I., & Sosa Bruchmann, E. C. (2016, May). Framework para data mining educativo: formalización y aplicaciones. In *XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina)*.

<sup>23</sup> Martínez, G. R. S., & Ocampo, C. A. (2009). Técnicas de clasificación y análisis de representación del conocimiento para problemas de diagnóstico. *Scientia et Technica*, 2(42).

neuronales y métodos estadísticos. Los árboles de clasificación o de decisión se caracterizan por su sencillez, su campo de acción abarca diversas áreas como: el diagnóstico médico, juegos, predicciones meteorológicas y control de calidad, este es el modelo de aprendizaje inductivo más utilizado (Martinez, 2009).

Para esta investigación se entenderá MD como el proceso de selección, procesamiento y agrupación de datos relevantes para su posterior investigación a través de la utilización de sistemas automatizados de análisis de datos masivos, con el fin de detectar información relevante sobre patrones de comportamiento, tanto internos como externos a la organización, para descubrir oportunidades de mejora en la cadena de procesos.

---

## 2.1 Historia

La MD surge junto a la creación de bases de datos digitales durante los años 90, evolucionando a partir del avance de las tecnologías y con la creación de herramientas de análisis capaces de procesar cantidades cada vez mayores de datos. Han, Pei y Kamber explican este proceso a través de la siguiente línea de tiempo:

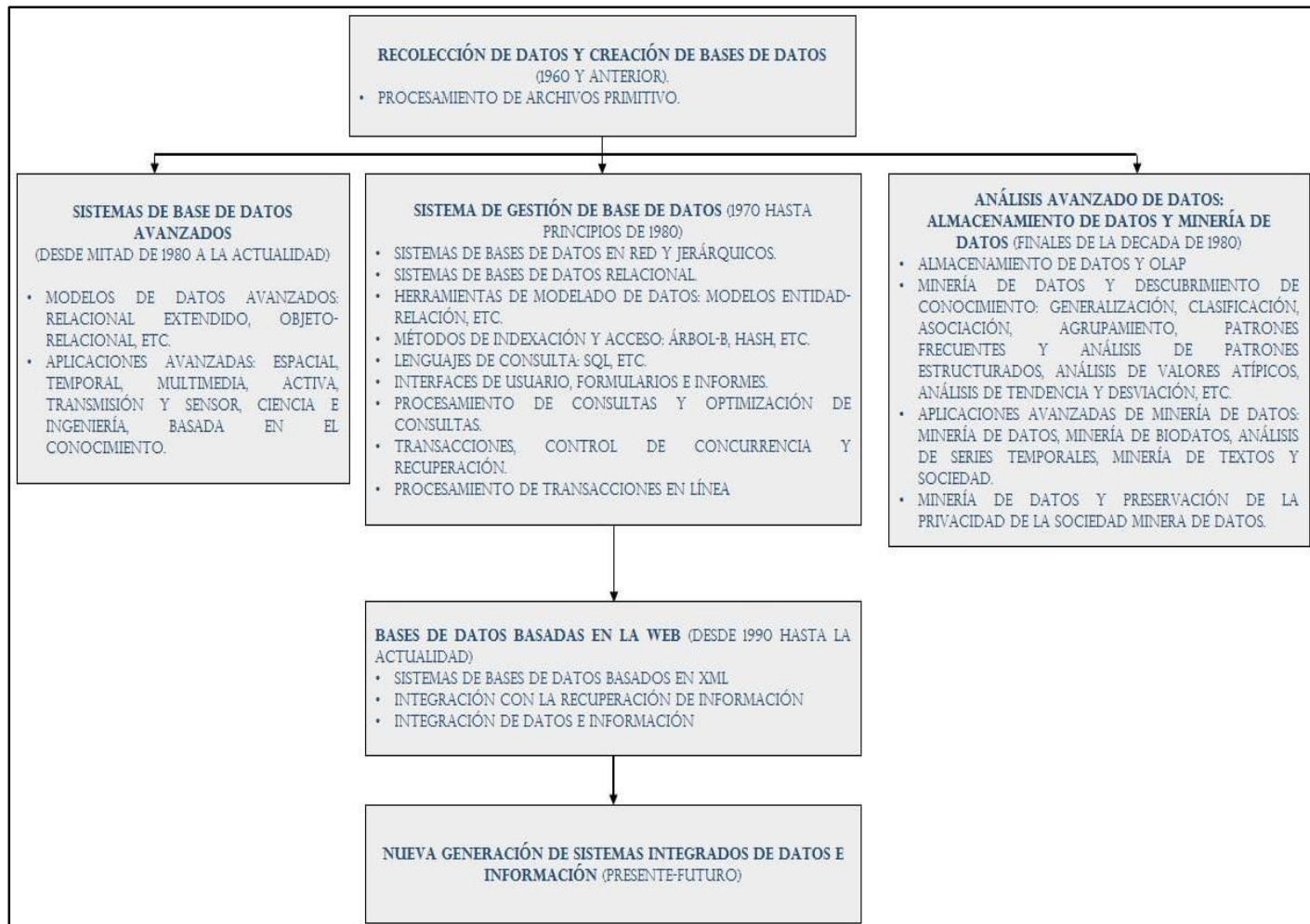


Fig. 2 La evolución tecnológica de los sistemas de bases de datos<sup>24</sup>.

La figura 2 representa la evolución de los sistemas de bases de datos desde su aparición en 1960 con la creación de archivos primitivos cuya función única era almacenar datos. Posteriormente se desarrollan modelos que utilicen los beneficios tecnológicos de la época: sistemas de bases de datos en red, interfaces, registros de transacciones, la posibilidad de hacer consultas entre

<sup>24</sup> Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier. La traducción es nuestra.

otros. En cuanto al análisis de datos avanzado, no surge hasta finales de 1980, donde algunas organizaciones utilizan los datos almacenados para detectar patrones frecuentes, análisis de valores atípicos, tendencia y desviación.

En la actualidad la mayoría de los sistemas trabaja en alguna proporción con una base de datos, que pueden ser más o menos complejas dependiendo de sus necesidades; algunas de ellas son compatibles con lenguajes web como xml, otras son capaces de integrarse con la recuperación de información y se ha vuelto habitual el contar con la opción para la creación de informes.

Ante este avance vertiginoso surge la pregunta: ¿Qué avances permitieron esta evolución en las bases de datos?

Se reconocen cuatro factores importantes que nos han llevado a este suceso:

1. El abaratamiento de los sistemas de almacenamiento tanto temporal como permanente.
2. El incremento de las velocidades de cómputo en los procesadores.
3. Las mejoras en la confiabilidad y aumento de la velocidad en la transmisión de datos.
4. El desarrollo de sistemas administradores de bases de datos más poderosos (Virsedá y Román, 2011)<sup>25</sup>

Virsedá y Román agregan antecedentes sobre el surgimiento de data mining: Ya desde los años sesenta los estadísticos manejaban términos como data fishing, data mining o data archaeology con la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con ruido. A principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros, empezaron a consolidar los términos de data mining y KDD. A finales de los años ochenta sólo existían un par de empresas dedicadas a esta tecnología; y en 2002 existían más de 100 empresas en el mundo que ofrecían alrededor de 300 soluciones. Las listas de discusión sobre este tema las forman investigadores de más de ochenta países (Virsedá y Román, 2011).

La popularidad de la minería de datos ha seguido creciendo durante la última década, con un énfasis particular en la extracción de datos no estándar.

## 2.2 Metodología

La MD para la obtención de información relevante, es posible gracias a la convergencia de tecnologías que permiten el tratamiento automatizado de datos, la gestión de bases de datos junto a los estudios métricos y estadísticos

---

<sup>25</sup> Virsedá, F. & Ramón, J. (2011). Minería de Datos y Aplicaciones. Disponible en: <<http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>> 4>.

de información <sup>26</sup>, concretando en métodos para descubrir conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información o base de datos. Entre las más utilizadas destacan:

### 2.2.1 KDD

Knowledge Discovery Databases o KDD, es una metodología que pretende obtener conocimiento a partir de datos almacenados en data warehouse (bodegas de datos). No es un proceso automático sino que es un proceso repetitivo que explora volúmenes muy grandes de datos de manera exhaustiva, para determinar relaciones. Al finalizar se extrae información de calidad que se puede usar para dibujar conclusiones basadas en relaciones o modelos dentro de los datos (Gutierrez y Molina, 2016)<sup>27</sup>.

El autor Gorbea-Portal (2013) reconoce las siguientes etapas:

1. **Selección de datos:** en esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde la o las fuentes de datos.
2. **Procesamiento:** limpieza de los datos para remover los datos inconsistentes que adicionan ruido. Integración de los datos, combinación de múltiples datos fuentes, en datos de mayor nivel de integración.
3. **Transformación:** transformación de los datos, consolidar los datos en formas más apropiadas para ser utilizados en la Minería de Datos.
4. **Minería de Datos:** Proceso esencial donde son aplicados métodos inteligentes para extraer patrones de los datos.
5. **Interpretación/ Evaluación:** identificar los patrones que verdaderamente representan el conocimiento basado en algunas mediciones interesantes, evaluar si estos corresponden a los datos en posesión, para finalmente presentar el conocimiento minado al usuario, mediante el uso de técnicas de visualización y representación de conocimiento.

---

<sup>26</sup> Gorbea-Portal, S. (2013). Tendencias transdisciplinarias en los estudios métricos de la información y su relación con la gestión de la información y del conocimiento. *Perspectivas em Gestão & Conhecimento*, 3(1), 13-27.

<sup>27</sup> Gutiérrez, J. A., & Molina, B. (2016). Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales. *Revista Ontare*, 3(2), 33-51.

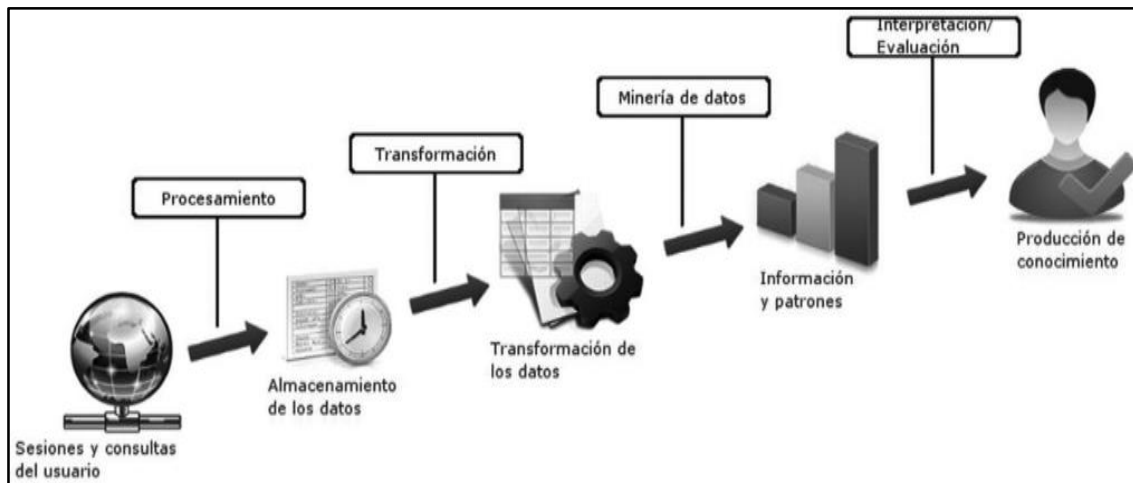


Fig. 3 Esquema de las etapas del proceso de descubrimiento del conocimiento en bases de datos [5].

### 2.2.2 CRISP-DM

Modelo para la minería de datos en el campo empresarial que pretende facilitar la realización de nuevos proyectos con características similares, optimizar la planificación y dirección de éstos, reducir su complejidad y permitir hacerle un mejor seguimiento a ellos (Gondar Nores, 2004)<sup>28</sup>.

La *guía paso a paso de data mining: CRISP-DM 1.0* (Clinton et al., 2000)<sup>29</sup> define los siguientes pasos como vitales para la implementación de la metodología CRISP-DM :

1. **Comprensión del negocio:** Esta fase se enfoca en entender los objetivos del proyecto y sus requerimientos desde una perspectiva de negocios, para descubrir desde un principio los factores que pueden influir en el resultado del proyecto. Una posible consecuencia de omitir este paso será el de invertir una gran cantidad de esfuerzo para encontrar la respuesta correcta a la pregunta equivocada.
2. **Comprensión de los datos:** La fase de comprensión de datos comienza con la recopilación de datos inicial y continúa con actividades que le permiten familiarizarse con los datos de la fase de Comprensión del negocio. Intenta identificar problemas de calidad en los datos,

<sup>28</sup> Gondar Nores, J.-E.(2004). Metodologías para la Realización de Proyectos de Data Mining. Disponible en: <<http://www.estadistico.com/arts.html?20040426>>.

<sup>29</sup> Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

descubrir los primeros datos sobre los datos y / o detectar subconjuntos interesantes para formar hipótesis con respecto a la información oculta.

3. **Preparación de los datos:** La fase de preparación de datos cubre todas las actividades necesarias para construir el “dataset” final [datos que se incorporarán a la(s) herramienta(s) de modelado] a partir de los datos brutos iniciales. Es probable que las tareas de preparación de datos se realicen varias veces y no en ningún orden prescrito. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para herramientas de modelado.
4. **Modelado:** En esta fase se seleccionan varias técnicas de modelado y calibrado a parámetros óptimos. Normalmente hay varias técnicas de MD para cada problema. Algunas tienen requerimientos específicos en la forma de expresar los datos. Aún así, regularmente es necesario volver al paso anterior y preparar los datos nuevamente.
5. **Evaluación:** en esta etapa del proyecto se construyen modelos de alta calidad de acuerdo a la perspectiva de análisis de datos. Antes de proceder a la implantación final del modelo, es importante evaluar el modelo exhaustivamente y revisar las etapas tomadas para construirlo, a modo de estar seguro que se cumplirá con los objetivos del proyecto esperados. Un objetivo clave es determinar si hay algún problema del negocio importante que no se haya tenido suficientemente en cuenta. Al final de esta fase, se debe alcanzar una decisión sobre el uso de los resultados de la minería de datos.
6. **Implantación:** en esta etapa se se presenta el conocimiento ganado de forma que sea utilizable por los clientes. A menudo implica la aplicación de modelos “en vivo” dentro de los procesos de la toma de decisiones en una organización, por ejemplo, la personalización en tiempo real de las páginas web. Según los requisitos, la fase de implementación puede ser tan simple como generar un informe o tan compleja como implementar un proceso repetible de minería de datos en toda la empresa. En muchos casos, es el cliente, no el analista de datos, quien lleva a cabo los pasos de implementación. Sin embargo, incluso si el analista llevará a cabo el esfuerzo de despliegue, es importante que el cliente comprenda por adelantado qué acciones deben llevarse a cabo para poder utilizar realmente los modelos creados.



Fig. 4 Modelo de proceso de la Metodología CRISP-DM (Díaz, 2016)<sup>30</sup>.

### 2.2.3 CMIN

Herramienta CASE integrada (aplicaciones informáticas destinadas a aumentar la productividad en el desarrollo de software) basada en la metodología CRISP-DM, que orienta el desarrollo de los proyectos, dando información detallada en cada paso de la minería de datos, facilita la integración de los procesos y el cumplimiento de estos en la ejecución del proyecto. Además su funcionalidad extensible, permite a la comunidad realizar modificaciones y mejoras, las que después de ser evaluadas y aprobadas son incluidas a la herramienta final (Cobos Et. Al, 2010)<sup>31</sup>.

Para comprender mejor el funcionamiento de la CMIN primero se presenta el modelo conceptual del sistema, con los principales conceptos y las relaciones existentes entre éstos:

1. **Usuarios:** comprende a las personas que pueden utilizar el sistema, los cuales pueden ser novatos o expertos en proyectos de minería de datos.
2. **Módulo de procesos:** es el que permite la gestión de procesos, entre ellos CRISP-DM. La definición de procesos representa la acción de registrar un proceso mediante la agregación y definición de sus pasos, campos o actividades que se proponen para el desarrollo de un proyecto de minería de datos. Los reportes son los documentos o

<sup>30</sup> Díaz, María Victoria. (2016). Metodología CRISP-DM – Parte 1. Disponible en: <<http://smartbasegroup.com/metodologia-crisp-dm-parte-i/>>.

<sup>31</sup> Cobos, C., Zuñiga, J., Guarín, J., León, E., & Mendoza, M. (2010). CMIN-herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *Ingeniería e Investigación*, 30(3).

entregables que se deben proveer durante un proyecto, y que son soporte de la ejecución de él.

3. **Procesos:** son los pasos que se han agregado a la CMIN y que sirven como base para gestionar los proyectos de minería con la herramienta.
4. **Módulo de proyectos:** representa el módulo de gestión de proyectos de minería de datos basado en uno de los pasos previamente adicionado en el módulo de procesos. Los proyectos comprenden el conjunto de procesos que se han creado en la CMIN y que están en curso o han sido terminados. Los campos o actividades de un paso son las tareas específicas que se deben realizar para cumplir con el objetivo del paso al que pertenecen. Los resultados representan los productos de la realización de una actividad, que pueden ser: una sugerencia, un texto explicativo o una plantilla de información que se debe diligenciar.
5. **Workflow (WF):** entorno gráfico que permite a los usuarios gestionar modelos de minería de datos basados en las tareas de minería definidas en la CMIN.
6. **Agregación dinámica de DLL:** es el módulo que permite la gestión de objetos (nuevos algoritmos) que sirven para la ejecución del WF, por medio de DLL. Los tipos de objetos del flujo de trabajo representan el conjunto de tipos de objetos reconocidos por la CMIN para ser agregados y posteriormente utilizados por el WF. Las interfaces abarcan el conjunto de contratos de software (por ejemplo, en clasificación, agrupación o reglas de asociación) que deben cumplir las DLL para agregarlas al conjunto de objetos que serán utilizados por el WF. Las DLL son el conjunto de algoritmos que posee actualmente la CMIN en su batería (objetos del WF).
7. **Objetos de WF:** comprenden el conjunto de objetos que se agregan a la CMIN y pueden utilizarse en el WF, el cual puede crecer a medida que los usuarios hagan nuevas implementaciones de cualquiera de los tipos de objetos del WF especificados en la CMIN.
8. **Servidor CMIN:** es el que aloja nuevas definiciones de procesos, así como nuevas implementaciones de objetos (algoritmos) del WF por medio de DLL para que los usuarios actualicen la CMIN si así lo requieren, ya que ella se ejecuta independientemente de este servidor.

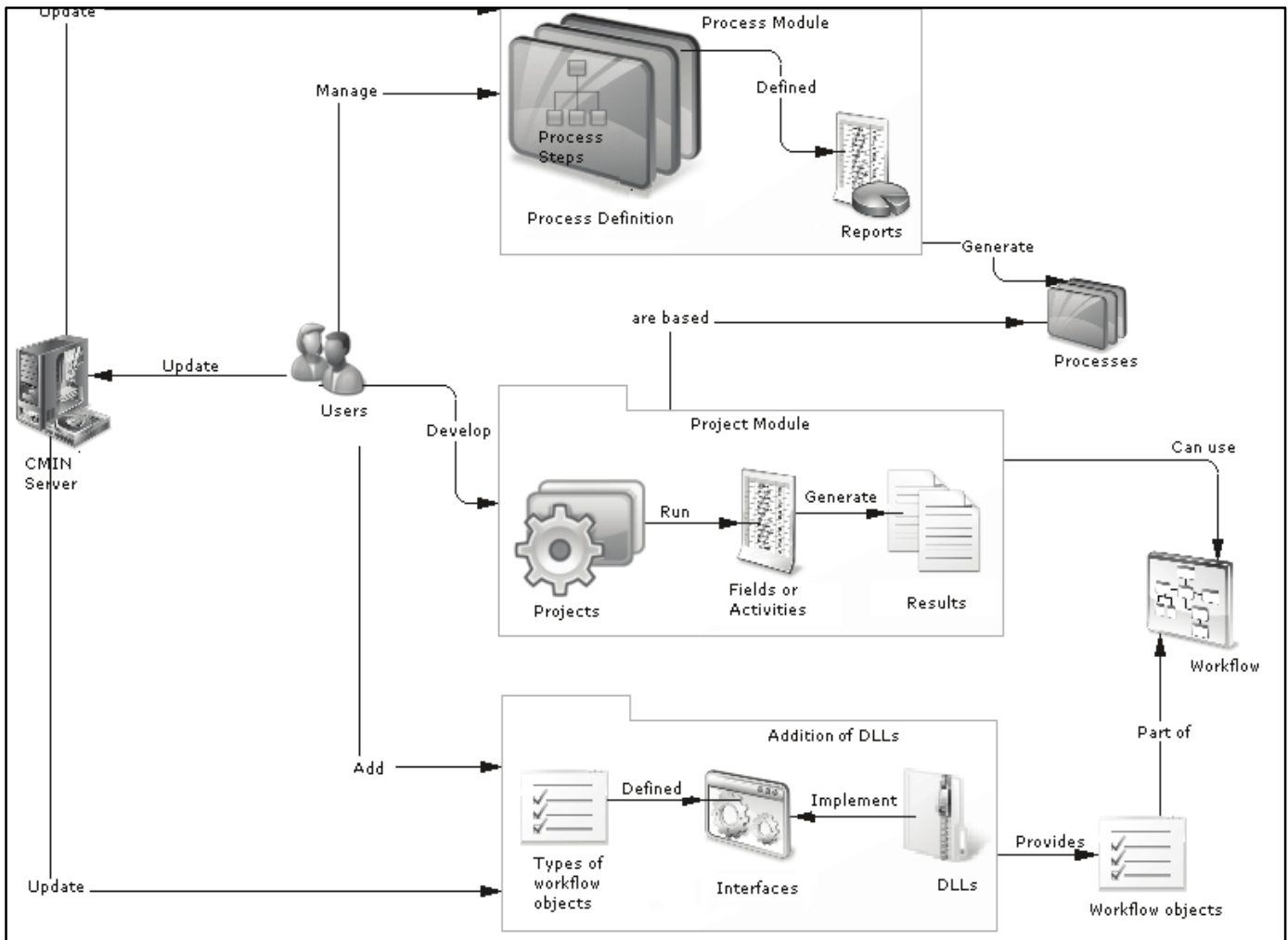


Fig. 5 Modelo conceptual de CMIN por Cobos (2010).

## 2.2.4 SEMMA

El Instituto SAS es el encargado de desarrollador este sistema, que si bien no es una metodología, es una herramienta desarrollada para aplicar variadas metodologías para la minería de datos. Se define como el proceso de selección, exploración y modelado de grandes cantidades de datos para descubrir patrones de negocio desconocidos. El nombre de esta terminología es el acrónimo correspondiente a las cinco fases básicas del proceso (Sample, Explore, Modify, Model, Assess) (2006)<sup>32</sup>, descritas de la siguiente manera:

- 0 (opcional) **Prueba:** se prueban los datos extrayendo una porción de un conjunto de datos significativa, lo suficientemente grande como para contener información pertinente, pero lo suficientemente pequeña como

<sup>32</sup> SAS Enterprise Miner. (2006). SEMMA. Disponible en: [http://faculty.smu.edu/tfomby/eco5385\\_eco6380/data/SPSS/SAS%20SEMMA.pdf](http://faculty.smu.edu/tfomby/eco5385_eco6380/data/SPSS/SAS%20SEMMA.pdf).

para ser manipulada con rapidez. Para un costo y rendimiento óptimos, SAS Institute propone una estrategia de muestreo que se aplique a una muestra confiable, que sea estadísticamente representativa de fuentes de datos ya detallados. Extraer una muestra representativa en lugar de todo el volumen reduce el tiempo de procesamiento requerido para obtener información crucial. Si aparecen patrones generales en los datos como un todo, estos serán reconocibles en una muestra representativa. Si un nicho es tan pequeño que no es representado en una muestra y, sin embargo, tan importante que influye en el panorama general, se puede descubrir utilizando métodos de resumen. También se aboga a crear conjuntos de datos particionados con el nodo Partición de datos:

- i. Entrenamiento: utilizado para ajustar el modelo.
- ii. Validación: se usa para la evaluación y para evitar el ajuste excesivo.
- iii. Prueba: se usa para obtener una evaluación honesta de qué tan bien se generaliza un modelo.

1. **Explorar:** explorar los datos buscando tendencias y anomalías imprevistas para obtener entendimiento e ideas. La exploración ayuda a refinar el proceso de descubrimiento. Si la exploración visual no revela tendencias claras, se pueden explorar los datos a través de técnicas estadísticas que incluyan análisis de factores, análisis de correspondencia y agrupamiento. Por ejemplo, en la minería de datos para una campaña de correo directo, la agrupación puede revelar grupos de clientes con distintos patrones de pedido. Conocer estos patrones crea oportunidades para correos personalizados o promociones.
2. **Modificar:** modificar los datos creando, seleccionando y transformando las variables para enfocar el proceso de selección del modelo. Según los descubrimientos en la fase de exploración, es posible que se necesite manipular ciertos datos para incluir información, como la agrupación de clientes y subgrupos significativos, o para introducir nuevas variables. Es posible que también se deban buscar valores atípicos y disminuir el número de variables para reducirlas a las más importantes. También es posible que se deban modificar los datos “minados” cuando cambien. Dado que la minería de datos es un proceso dinámico e iterativo, se pueden actualizar los métodos o modelos de minería de datos cuando haya nueva información disponible.
3. **Modelar:** modelar sus datos permite que el software busque automáticamente una combinación de datos que permita predecir de

manera fiable el resultado deseado. Las técnicas de modelado en la minería de datos incluyen redes neuronales, modelos basados en árboles, modelos logísticos y otros modelos estadísticos, como el análisis de series de tiempo, el razonamiento basado en la memoria y los componentes principales. Cada tipo de modelo tiene fortalezas particulares y es apropiado dentro de situaciones específicas de minería de datos según los datos. Por ejemplo, las redes neuronales son muy buenas para encajar relaciones no lineales altamente complejas.

4. **Evaluar:** evaluar los datos estimando la utilidad y confiabilidad de los hallazgos del proceso de minería de datos y calcular qué tan bien funciona. Un medio común para evaluar un modelo es aplicarlo a una parte de los datos reservados durante la etapa de muestreo (muestra). Si el modelo es válido, debería funcionar para esta muestra reservada, así como para la muestra utilizada para construir el modelo. Del mismo modo, se puede probar el modelo con datos conocidos. Por ejemplo, si se sabe qué clientes en un archivo tienen altas tasas de retención y el modelo predice la retención, se puede verificar si el modelo selecciona estos clientes con precisión.

Al evaluar los resultados obtenidos de cada etapa del proceso SEMMA, se puede determinar cómo modelar las nuevas preguntas planteadas por los resultados anteriores, y por lo tanto volver a la fase de exploración para un refinamiento adicional de los datos.

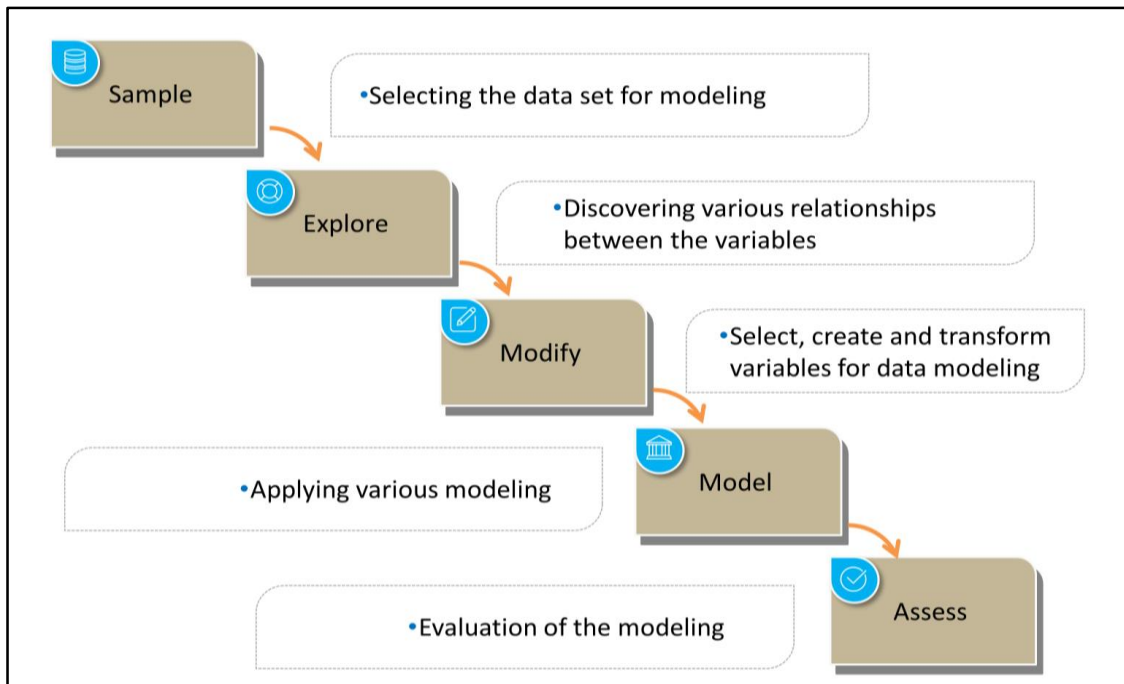


Fig. 6 Metodología SEMMA<sup>33</sup>

## 2.3 Tecnologías

### 2.3.1 Programas

En el contexto de la MD existen una serie de programas que permiten llevar a cabo la tarea de extraer, cargar, interpretar y generar nueva información a través de las predicciones. Entre los más utilizados se encuentran Oracle, Alteryx, Rapid Miner y SAS. Cada uno posee diversas características que se ajustan a las capacidades de los usuarios.

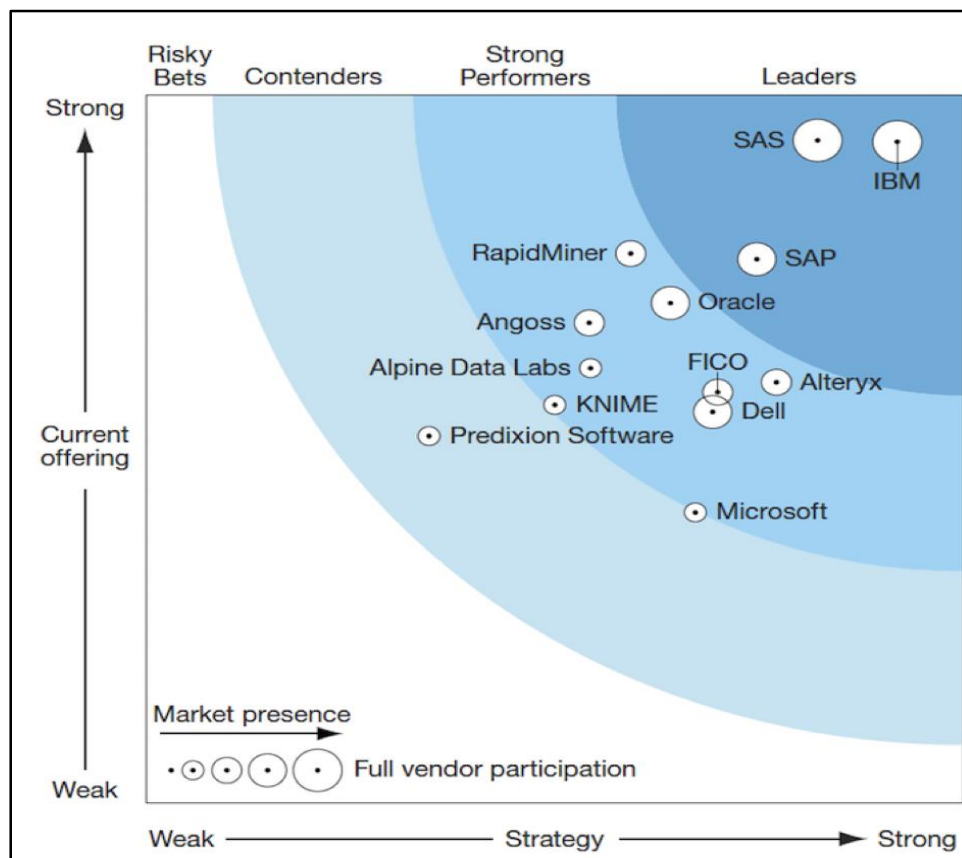


Fig. 7 El ranking de las mejores soluciones de análisis predictivo para empresas (Cía, 2015)<sup>34</sup>

<sup>33</sup> YOUR FREE TEMPLATES. (2017). *Phases of SEMMA*. Disponible en: <https://yourfreetemplates.com/free-data-mining-template/>.

<sup>34</sup> Cía, Juan F. (2015). El Ranking de las mejores soluciones de análisis predictivo para empresas. Disponible en: <https://bbvaopen4u.com/es/actualidad/el-ranking-de-las-mejores-soluciones-de-analisis-predictivo-para-empresas>

En esta oportunidad se analizarán los dos últimos que se mencionan. En primer lugar se presentará una breve definición de cada uno, seguido de una tabla comparativa.

### 2.3.1.1 Rapid Miner

Es un programa de características comerciales que cuenta con una versión gratuita. Permite la realización de minería de datos. Su principal virtud es que no se requieren mayores conocimientos informáticos para ejecutarlo dada su interfaz gráfica compuesta por operadores, lo que unidos forman un modelo. Su primera versión fue gestada por la Universidad de Dortmund, Alemania, en el año 2001, cuenta con licencia GPL y está hospedado en SourceForge desde el año 2004.<sup>35</sup>

Rapid Miner es muy sencillo de configurar y administrar, lo que facilita su uso.

Entre sus principales características se encuentran:

Características Rapid Miner		
1. Es multiplataforma.	4. Es extensible.	7. Puede usarse de distintas maneras:
2. Está creado en Java script.	5. Posee gráficos y heramientas que permiten el muestreo de datos.	- En batch.
3. Permite generar programas usando lenguaje script.	6. Representación interna de procesos de análisis de datos de ficheros XML.	- Líneas de comando.
		- A través de GUI.
		- Otros programas, por medio de llamadas a sus bibliotecas.

Fig.8 Características de Rapid Miner<sup>36</sup>

<sup>35</sup> Beltrán, David. Poveda, Diego. Rapid Miner. (2010).

<sup>36</sup> Elaboración propia. Basado en Rapid Miner. Beltrán, David y Poveda, Diego. (2010).

Actualmente, la comunidad de Rapid Miner cuenta con 377.483 miembros, 268 online y 45.320 post en su página y fue nombrado el líder del año 2018 en el Gartner Magic Quadrant for Data Science and Machine Learning Platform<sup>37</sup>.

### 2.3.1.2 SAS

Al igual que rapid miner, SAS permite hacer las mismas tareas, con la salvedad de no contar con una interfaz gráfica de operadores y modelos, lo que lo hace menos amigable. Este programa requiere contar con mayores habilidades estadísticas y/o informáticas para ejecutarlo. Sin embargo, es un programa de alta presencia en el mercado y esto hace que los usuarios confíen en él pues tienen la experiencia de otros a quienes les ha funcionado.

Entonces, se abre el cuestionamiento, ¿Qué programa es mejor entre éstos dos? Todo dependerá del tipo de usuario, sus habilidades y para qué se necesita el programa. Si bien es cierto todos realizan predicciones en base a una serie de datos que se deben cargar, cada uno posee distintas características. Por ejemplo, Rapid Miner dado lo intuitivo que es su manejo y configuración además de sus operadores y modelos, lo hacen más útil para pequeños usuarios sin mayores conocimientos informáticos.

Sas, por otra parte, es muy popular entre usuarios como grandes empresas, cuyos requerimientos son distintos a los de pequeños grupos de usuarios (G2 CROWD, 2018)<sup>38</sup>.

Cabe destacar que Sas, como la mayoría de los softwares, está compuesto por varias versiones. Para MD se suelen emplear Sas Visual Statistics y Sas Analytics Visual Explorer (Cía, 2015).

A continuación se presentará la comparación entre un programa y otro (Rapid Miner y Sas).

---

<sup>37</sup> RapidMiner Community (2018). Disponible en: <<https://community.rapidminer.com/>>

<sup>38</sup> G2 CROWD. (2018). *Compare RapidMiner vs SAS*. Disponible en: <<https://www.g2crowd.com/compare/rapidminer-studio-vs-sas-business-intelligence>>.

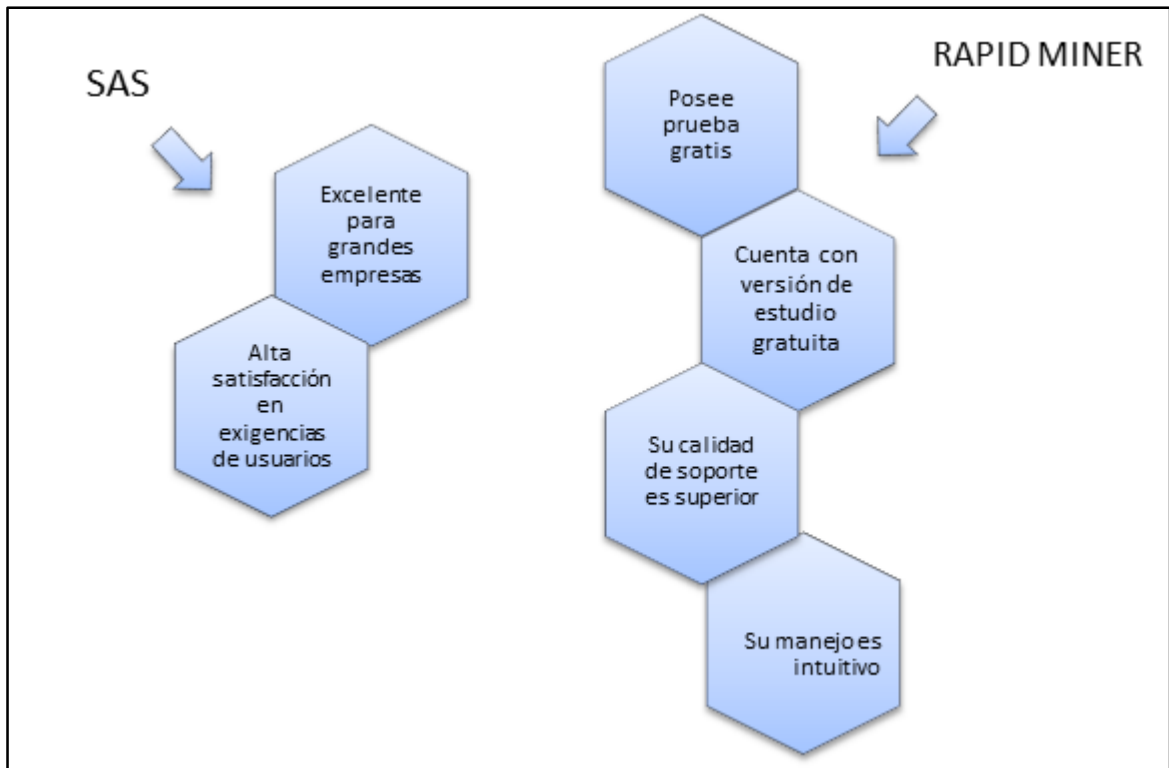


Fig. 9 Rapid Miner VS SAS<sup>39</sup>

### 2.3.2 Lenguajes

En el mundo de la informática, existe una serie de lenguajes propios de la programación que permiten precisamente eso, programar. Transforman el lenguaje común y los datos para que éstos puedan ser interpretados por un software.

En esta oportunidad se hablará de Python y R. Ambos lenguajes de programación tienen múltiples usos, entre ellos se encuentra la MD, de ahí la importancia para esta investigación.

#### 2.3.1.1 R

R es un lenguaje de programación desarrollado por Ross Ihaka y Robert Gentleman en el año 1996. Este lenguaje posee una especial característica. Para funcionar debe valerse de un determinado entorno, el cual le permitirá posteriormente ser leído por el computador<sup>40</sup>.

<sup>39</sup> Elaboración propia. Basado en G2 CROWD. Compare RapidMiner vs SAS. Disponible en: <https://www.g2crowd.com/compare/rapidminer-studio-vs-sas-business-intelligence>

<sup>40</sup> Paradis, Emmanuel. R para Principiantes.(2003).

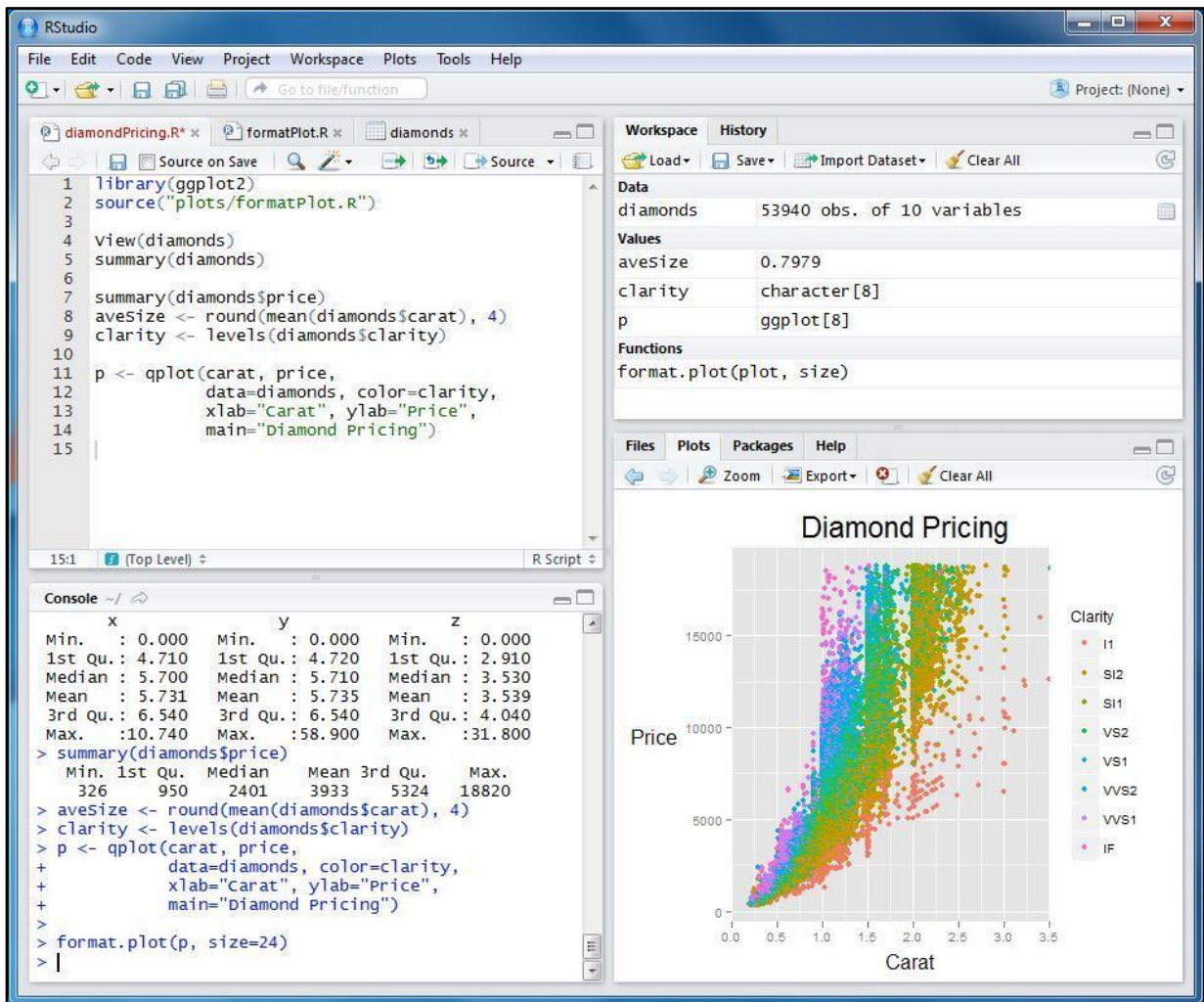


Fig. 10 R Studio. Elaboración propia.

En esta imagen es posible visualizar R como lenguaje de programación (parte izquierda de la pantalla) y su visualización gracias al entorno R Studio, un programa que permite leer e interpretar los código de R (parte derecha de la pantalla).

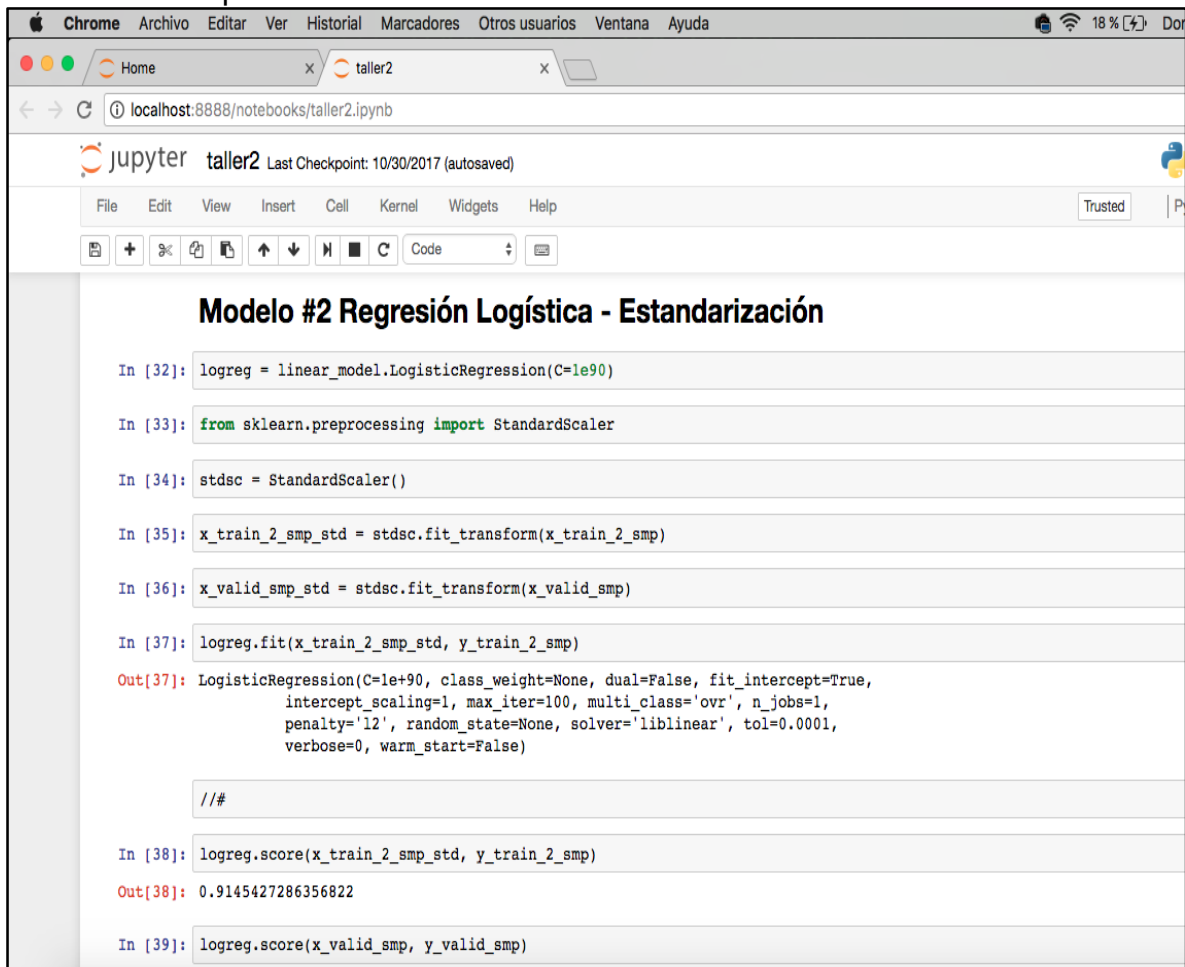
R permite manipular cálculos, datos y gráficos. Es un lenguaje muy dinámico y posee diversas versiones, las cuales no siempre son compatibles entre ellas.<sup>41</sup>

<sup>41</sup> R Development Core Team. Introducción a R. (2000).

### 2.3.2.2 Python

Python es un rápido y poderoso lenguaje de programación que permite integrar sistemas de manera efectiva. Python posee diversos usos, entre ellos destacan el desarrollo web, la enseñanza de la programación misma, el mundo de los negocios, informática numérica y científica, entre otros<sup>42</sup>. (La traducción es nuestra).

Al igual que R para que Python pueda ser leído por un computador, necesita de un entorno que lo sustente.



```
Modelo #2 Regresión Logística - Estandarización

In [32]: logreg = linear_model.LogisticRegression(C=1e90)

In [33]: from sklearn.preprocessing import StandardScaler

In [34]: stdsc = StandardScaler()

In [35]: x_train_2_smp_std = stdsc.fit_transform(x_train_2_smp)

In [36]: x_valid_smp_std = stdsc.fit_transform(x_valid_smp)

In [37]: logreg.fit(x_train_2_smp_std, y_train_2_smp)

Out[37]: LogisticRegression(C=1e+90, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
verbose=0, warm_start=False)

//#

In [38]: logreg.score(x_train_2_smp_std, y_train_2_smp)

Out[38]: 0.9145427286356822

In [39]: logreg.score(x_valid_smp, y_valid_smp)
```

Fig. 10 Python en Jupyter. Elaboración propia.

En esta imagen, se visualizan una serie de códigos, los cuales aún no se han interpretado.

<sup>42</sup> Python Software Foundation. Disponible en: [<https://www.python.org/about/>](https://www.python.org/about/)

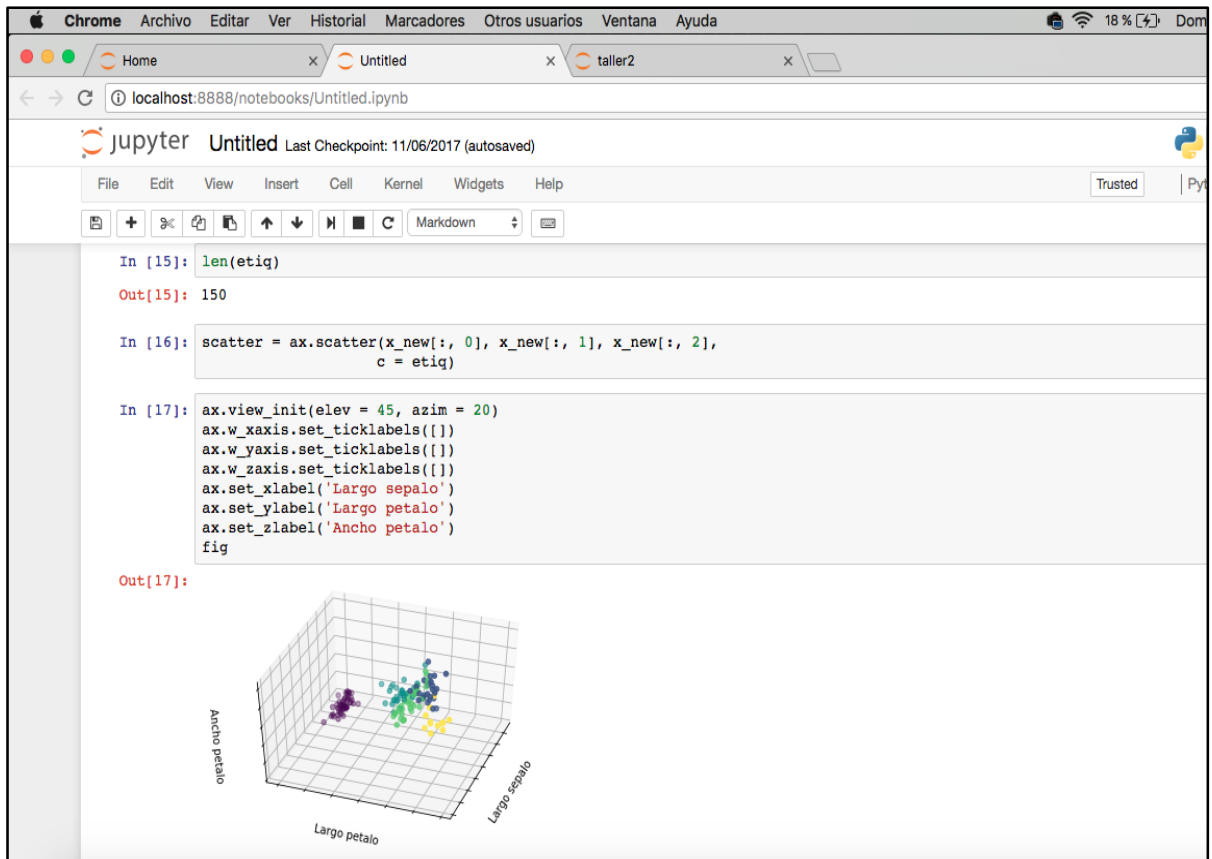


Fig. 11. Python en Jupyter. Elaboración propia.

A diferencia de la imagen anterior, ésta permite visualizar la interpretación del código y los interpreta a un diagrama de puntos.

Como se mencionó anteriormente, Python se vale de un entorno para reflejar los códigos y que éstos se entiendan más fácilmente. Los códigos de las imágenes (números) se desarrollaron en un entorno llamado Jupyter.

## 2.4 Usos

### Detectar fraudes financieros

Data mining puede ser utilizado para detectar fraudes financieros dentro de las organizaciones, los que se encuentran divididos en dos categorías: fraude en los estados financieros y fraude en las transacciones (Jans et al., 2011) <sup>43</sup>. Fraude en los estados financieros se define como "la declaración errónea intencional de ciertos valores financieros para mejorar la apariencia de

<sup>43</sup> Jans, M., Van Der Werf, J. M., Lybaert, N., & Vanhoof, K. (2011). A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications*, 38(10), 13351-13359.

rentabilidad y engañar a los accionistas o acreedores", mientras que el fraude de transacciones captura el proceso de robo a los activos de la organización.

Las herramientas de Minería de datos enfocadas en la detección de fraudes financieros han aparecido en escena para abordar este problema y proporcionar soluciones confiables para las empresas. El fraude financiero normalmente se descubre a través del proceso de detección de valores atípicos habilitado por las técnicas de minería de datos, que también identifican información valiosa al revelar tendencias ocultas, relaciones, patrones encontrados en una gran base de datos (Albashrawi, 2016)<sup>44</sup>.

Además de la detección de fraudes, Albashrawi detecta un amplio rango de aplicaciones en el área de las organizaciones, por ejemplo, predicción de bancarrota, previsión de ventas y optimización de la programación empresarial.

### **Aplicaciones de Análisis de Información Textual: Corpus Lingüísticos**

Es posible crear herramientas utilizando minería de datos, para construir corpus lingüísticos, que pueden ser utilizados como material de entrenamiento para sistemas de minería de datos y de extracción de información, en especial sobre texto no estructurado (Castillo, 2017)<sup>45</sup>.

Mediante el proyecto ADTD desarrollado por el departamento de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba, se propone abordar la detección de similitudes entre diferentes textos. Por ejemplo se puede clasificar si dos fragmentos de textos se relacionan entre sí, y en ese caso, qué grado de relación presentan entre ellos.

A efectos de poder abordar este tipo de problemas, es necesario contar con conjunto de datos (material lingüístico) apropiado para poder construir algoritmos de aprendizaje sobre los mismos. Entre los usos que tiene este proyecto se encuentran:

- Proveer un mecanismo de clasificación de pares de texto con paráfrasis y facilitar la lectura y estudio de otros corpus.
- Contar con una herramienta que permita automatizar la clasificación de fragmentos de textos en base a la presencia o ausencia de ciertos

---

<sup>44</sup> Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science*, 14(3), 553-569.

<sup>45</sup> Castillo, J. J., & Cardenas, M. E. (2017, August). *Aplicaciones de análisis de información textual: corpus lingüísticos*. In XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires).

fenómenos lingüísticos que se mantengan en el mismo texto o que se mantenga entre dos fragmentos de textos.

- Generación de corpus etiquetado. El corpus permite incrementar la capacidad de reconocimiento y de clasificación de los algoritmos utilizados [45].

### **En el área de salud**

Analizar características de enfermedades y crear perfiles de las mismas, que empaten con los síntomas de un paciente es una tarea que requiere herramientas tecnológicas, como MD para analizar grandes conjuntos de datos heterogéneos (Reyes & Thompson & Vanoye & Penna, 2018)<sup>46</sup>.

Al realizar un análisis de los registros de salud como a los historiales de salud de pacientes, fechas de ingreso a hospitales, enfermedades adquiridas, padecimientos, tratamientos a los que se ha sometido, notas médicas, resultados clínicos, entre otros datos es posible obtener información que ayude a comprender el comportamiento de cierta enfermedad y cómo se ha tratado, contribuyendo a encontrar patrones de comportamiento de enfermedades, así como posibles soluciones para su tratamiento [46].

### **Predecir el rendimiento deportivo**

Como ha podido apreciarse, la minería de datos ofrece grandes posibilidades para el análisis de datos, permitiendo el descubrimiento de patrones y regularidades de alta significación. Aunque la minería de datos ha sido aplicada fundamentalmente en el marco empresarial, su potencialidad en la esfera deportiva es prominente.

Pagola, Pinheiro y Despachelles (2010) innovan utilizando la minería de datos para detectar patrones de rendimiento, entre deportistas de voleibol, detectando los factores emocionales que influyen en el desempeño de jugadores, desechando características que se pensaban como positivas para los jugadores<sup>47</sup>.

A partir de este estudio pudo apreciarse, en el análisis descriptivo, los comportamientos de algunas variables psicológicas, descubriendo nuevas variables significativas o, incluso, descartando otras tradicionalmente consideradas como tal para el rendimiento de los atletas. Algunos ejemplos de las tendencias encontradas [47]:

---

<sup>46</sup> Reyes, G. V., Thompson, E. B., Vanoye, J. A. R., & Penna, A. F. (2018). MODELOS DE TECNOLOGIAS DEL BIG DATA ANALYTICS Y SU APLICACIÓN EN SALUD. *Pistas Educativas*, 39(128).

<sup>47</sup> Pagola, J. M., Pinheiro, M. S., & Despachelles, C. A. (2010). Evaluación del Rendimiento de los Voleibolistas mediante Minería de Datos. *Ingeniería Industrial*, 26(2), 6.

- Seis atletas con el temperamento colérico alto, la agresividad media y la ansiedad no alta tuvieron muy alto rendimiento. •
- Diez atletas con el temperamento colérico bajo tuvieron alto rendimiento. •
- La ausencia de la variable cooperación en las fases eliminatorias de un torneo, acrecentando su presencia en las fases finales.

En general, el estudio realizado permitió la evaluación de elementos metodológicos que pueden orientar a los expertos del deporte en la búsqueda e investigación de determinados factores que inciden en el rendimiento de los atletas, mostrando la posibilidad de trabajar en esta línea. Como resultado de este estudio experimental, los especialistas y directivos deportivos han comprendido la necesidad y conveniencia de aplicar estos resultados, promoviendo un estudio más profundo en esta área.

## 2.5 Aplicación en Chile

En Chile se han realizado escasas investigaciones vinculadas con la aplicación de minería de datos. Entre ellas se utiliza para **determinar el comportamiento de estudiantes durante su formación académica**, esperando predecir factores que afecten el rendimiento a lo largo de su vida estudiantil, con el fin de permitir a las unidades educativas distribuir de forma eficiente sus recursos.

En 2012 Angúlo presentó una investigación sobre los principales factores que afectan la deserción universitaria<sup>48</sup>, utilizando datos proporcionados por la Universidad de las Américas, específicamente de la Escuela de Tecnologías de la Información, abarcando las sedes Providencia, Santiago Centro y La Florida.

Para la implementación de este proyecto se adoptó la metodología CRISP-DM y se aplicaron los modelos de Redes Neuronales, Árboles de decisión y Cluster K-medianas para analizar el comportamiento de los estudiantes, evaluando factores como el puntaje promedio obtenido en la Prueba de Selección Universitaria (PSU), el promedio de notas obtenido en la enseñanza media, la edad a la fecha de Ingreso a la institución y el género de los estudiantes.

La exactitud de los modelos fue calculada a partir del conjunto de datos de pruebas, los cuales indicaron que ningún modelo predictivo fue capaz de arrojar resultados positivos, debido a esto se analizó el proceso y se llegó a la conclusión de que es muy probable que los datos de entrada no eran

---

<sup>48</sup> Angúlo, F., & Sergio, E. (2012). Modelo para la automatización del proceso de determinación de riesgo de deserción en alumnos universitarios.

suficientemente confiables. En general las instituciones no recolectan la suficiente información referente a la caracterización del estudiante al momento de ingresar a estudiar como para establecer modelos de predicción de retención asertivos.

Como recomendación, Ángulo presenta una lista de variables a considerar y recopilar, para la realización de estudios futuras:

Tipo	Atributo	Descripción	Posibles valores
<b>Individuales</b>	RUT	Identificador de la tupla	Los numeros de rut
	Nombre	Nombre y apellidos del estudiante	Todos los nombre posibles
	FecNac	Fecha de nacimiento del estudiante	rango de fechas posibles
	Sexo	Género del estudiante	M (masculino), F (femenino)
	Comuna	Comuna de residencia del estudiante	Las comunas de la region Metropolitana
	LugarDeVivienda	indicador si el estudiante vive con la familia, solo, o con otros	ConFamilia, con Amigos, solo
	PracticaDeporte	Indicador si el estudiante practica deporte con Regularidad	Si, No
	HabitosConsumo	Habitos de Consumo	Fuma, Bebe, ambos, no aplica
<b>Académicos</b>	FeIngreso	Fecha Ingreso a la Universidad	Fecha valida en el rango de estudio
	Prom_PSU	Promedio PSU	rango de puntages 200 a 900
	Prom_NEM	Promedio NEM	Rango de notas 4 a 7
<b>Institucionales</b>	SedeUniversidad	Sede donde estudia el alumno	LF, SC, PR, MP, CO, VL
	ComunaUniversidad	Comuna de la Sede de la universidad	Las comunas de la region Metropolitana
	Colegio	Colegio Origen	Nombre de los colegios
	ComunaColegio	Comuna Colegio	comunas de Chile
	TipoColegio	Tipo de administración del Colegio	Municipal, Particular Subencionado, Particular
	TipoEnseñanza	Tipo de enseñanza del Colegio	Científico-Humanista (EMCH), Técnico-
	PuntajeSIMCEColegio	Promedio SIMCE	resultados prueba SIMCE Colegio
<b>Socio económicos</b>	GrupoSocioeconomicoFamiliar	Nivel socioeconómico	Bajo, Medio Bajo, Medio, Medio alto, Alto
	NeduMadre	Nivel educativo de la Madre	Primarios o Sin estudios, Secundarios,
	NeduPadre	Nivel educativo de la Padre	Primarios o Sin estudios, Secundarios,
	TrabE	Trabajo del estudiante	Si, No
	<i>Estado</i>	El estado a predecir (La clase)	Titulado, Desertor

Fig. 12 Variables a considerar en estudios futuros (Angúlo, 2012).

### 3. Bibliomining

El término bibliomining fue acuñado por primera vez el año 2003 por los autores Scott Nicholson y Jeffrey Stanton. Es definido como el uso de la extracción de datos para examinar los registros de datos de una biblioteca <sup>49</sup> Es decir, es la unión entre la Bibliometría y las bases de datos para generar reportes en el marco de la información bibliotecaria para crear información nueva que posteriormente será usada para tomar decisiones. (Valeria Haro, Wilson Pérez, Víctor Saquicela 2016).

Dada la importancia de tomar buenas decisiones, el uso de la minería de datos en las bibliotecas (en este caso bibliotecas escolares CRA) se traduciría en una optimización de tiempos y espacios. Mejor manejo económico y de personal, es decir, bibliomining vendría a facilitar todos los procesos dentro de una biblioteca, transformándola en un espacio práctico y verdaderamente útil.

Bibliomining es un proceso que cuenta con seis fases o etapas que van desde el simple hecho de identificar el problemas hasta el resultado final o validación de resultados.

Identificar Problema	Este es el primer paso en las seis fases de acción aplicadas en Bibliomining. En este punto se busca identificar el área específica en que se trabajará.
Establecer fuentes de datos	Se deben obtener fuentes de datos tanto internas (datos propios que se encuentran dentro de la biblioteca, por ejemplo ordenamiento de estanterías, funcionamiento del sitio web, misión de visión de la institución, entre otros.) como externas (información foránea a la biblioteca, por ejemplo, ubicación geográfica de la unidad de información y perfiles de

---

<sup>49</sup> Nicholson, Scott. Stanton, Jeffrey. (2003). Gaining Strategic Advantage Through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries. Disponible en: [https://surface.syr.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1112&context=istp\\_ub](https://surface.syr.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1112&context=istp_ub)

	usuarios registrados en la base de datos).
Data Warehouse	Posterior a la identificación del problema y la recolección de datos, se procede a seleccionar sólo la información útil, es decir, se procede a limpiar los datos recolectados.
Análisis de datos	Con los datos ya seleccionados, se procede al análisis. Para ello se debe contar con una herramienta apropiada.
Patrones y reportes	Con la utilización de minería de datos, se investigan patrones de comportamiento, lo que posteriormente permitirá generar reportes
Resultados	Finalmente, se debe testear que los patrones recopilados arrojen resultados altos en porcentaje, para así validar todo el proceso.

Las 6 fases de Bibliomining<sup>50</sup>. Elaboración propia.

### 3.2 Usos

Los usos de Bibliomining son variados, pero generalmente están ligados al área de aprendizaje.

#### **Creación de colecciones automáticamente**

Bibliomining provee de un conjunto de herramientas que pueden ser utilizadas para descubrir patrones en largas cantidades de datos sin procesar, e

---

<sup>50</sup>Tabla basa en: Haro, Pérez, Saquicela. (2016). *Bibliomining para descubrir reglas de asociación en el Centro de Documentación Regional "Juan Bautista Vázquez" y Nicholson, Stanton.* y [47].

identificar los patrones necesarios para la creación de modelos para la ayuda automatizada de colecciones (Nicholson, 2003)<sup>51</sup>.

Bajo esta concepción es posible clasificar sitios web y documentos electrónicos con un alto índice de precisión (95%), siempre que cumplan ciertas características, como contar con acceso al texto completo y una metadata que se ajuste a 5 criterios identificado por Nicholson y un grupo de 42 bibliotecarios: criterio de autor, criterio de contenido, criterio organizacional, criterio de productor/medio y criterios externos (2003).

### **Descubrimiento de patrones de comportamiento de usuarios:**

Otra aplicación de bibliomining en un centro de documentación o biblioteca es permitir a los administradores obtener información acerca de los intereses que tienen los usuarios, y focalizar los recursos para ofrecer un mejor servicio mediante la adecuada toma de decisiones (Pérez et al., 2017)<sup>52</sup>.

Para esto se extrae la información obtenida de las bases de datos digitales relacionada con el proceso de búsquedas de recursos y préstamo de recursos. Posteriormente se utiliza alguna herramienta de análisis de datos, como R, para descubrir patrones de comportamiento a modo de generar nuevo conocimiento (Pérez et al., 2017).

Como resultado es posible descubrir los intereses de los usuarios en cuanto a áreas temáticas específicas, por ejemplo, se puede llegar a la conclusión de que se correlacionan las búsquedas y los préstamos sobre libros de química general en un centro de documentación, conocimiento que incentiva al bibliotecario a invertir recursos en esa área de interés.

### **La actualización de procesos**

La agilidad en los procesos de gestión de la biblioteca y en la respuesta a solicitudes de información, específicamente para informes como los exigidos por el Ministerio de Educación Nacional en los trámites de obtención de registros calificados y de acreditación de programas académicos. La capacidad de extracción sencilla y ordenada de datos permite a la organización facilitar la recopilación de datos importantes, con el fin de cumplir con procesos de fiscalización y acreditación [4].

---

<sup>51</sup> Nicholson, S. (2003). Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web- based scholarly research works. *Journal of the Association for Information Science and Technology*, 54(12), 1081-1090.

<sup>52</sup> Pérez, W., Haro, V., & Saquicela, V. (2017). *Bibliomining: Aplicación de text mining para descubrir preferencias de usuarios en el Centro de Documentación Regional "Juan Bautista Vázquez"*. *Maskana*, 8, 135-144.

### **Gestión económica: selección y adquisición de fondos**

En la gestión de las bibliotecas, como en cualquier organización, la toma de decisiones desde la dirección es una actividad fundamental. Especialmente delicada es en la

En la gestión de bibliotecas es importante la forma de asignar y utilizar los recursos económicos, y más teniendo en cuenta que la mayor parte de las bibliotecas tienen recursos limitados [6].

Así, la utilización de datos sobre el comportamiento de los usuarios en el uso de la colección resulta de gran ayuda e importancia para la toma de decisiones respecto a la selección y adquisición de fondos. Es decir, conocer qué tipo de material se usa más y cuáles de los ítems más solicitados cuentan con pocos ejemplares, resulta crucial y muy valioso para orientar futuras adquisiciones [6].

### **Capítulo 3: Propuesta de implementación.**

Antes de comenzar con la definición de criterios y posterior propuesta de implementación, es importante destacar que para este ejercicio se considerará la realidad habitual de la biblioteca escolar CRA, que cuenta con recursos limitados y personal sin habilidades avanzadas en el área de la informática, para que la metodología a proponer esté al alcance de todos y, en caso de contar con mayores recursos económicos o con un personal que posea mayores estudios en la materia, puedan utilizarlo también sin dificultad.

Entonces, nuevamente se hará mención a los principales desafíos que tienen los CRA, para luego darle paso a las posibles soluciones que ofrece el bibliomining.

Como primer ejemplo de desafío, se presenta la necesidad imperiosa de profesionalizar la red de coordinadores y encargados de las bibliotecas escolares CRA a lo largo de Chile. Esto implica y conlleva a la unificación de criterios a nivel país en cuanto a una estructuración clara y concisa de los profesionales a cargo de la biblioteca escolar CRA.

No hay que olvidar que, al tratarse de una biblioteca escolar, el principal usuario son menores de edad en plena formación, tanto académica como personal.

El bibliomining, como ya se había mencionado, es un buen elemento predictivo que intentará, desde un punto muy basal y pretendiendo ofrecer soluciones a este desafío, captar a los profesionales de la información cuya vocación esté ligada al área de las bibliotecas escolares CRA, al tiempo que, como agilizador de procesos que es, identificar los mayores déficits y/o deficiencias existentes en la profesionalización de los encargados de estas bibliotecas, para dar pie a su posterior mejora.

Además de una normativa a nivel nacional que regule la formación de los profesionales a cargo de una biblioteca escolar CRA, urge cumplir con otros desafíos latentes, tales como la necesidad de otorgar recursos igualitarios a todas las bibliotecas escolares CRA del país, independiente del nivel socioeconómico en que se desenvuelve el colegio. Se sabe que esto no es de solución fácil, por lo que en este punto surge la necesidad de aprender a manejar los recursos de manera óptima. Para ello es fundamental contar con profesionales que posean una formación similar y dentro de sus habilidades sepan manejar adecuadamente los recursos económicos para brindar el mejor servicio posible.

Como último desafío, se presenta la necesidad de fortalecer el uso de los recursos del aprendizaje en las bibliotecas escolares CRA que se mencionan

en el capítulo 2 podrían ser monitoreados, reforzados y/o replanteados con el correcto apoyo de bibliomining.

Bibliomining se presenta como un agilizador de procesos administrativos, permitiendo mejorar la toma de decisiones, por lo que a largo plazo sería de gran ayuda para resolver éstos y futuros desafíos que se vayan presentando a medida que evolucionan las políticas nacionales, las bibliotecas escolares CRA y sus usuarios.

### 3. Definición de criterios

Luego de la investigación realizada y de los datos recopilados, se definirán los criterios a aplicar al momento de recomendar una metodología para bibliotecas escolares CRA, la cual incluirá la selección del programa y tecnología más adecuado de MD en este caso particular. Los criterios a considerar serán los siguientes:

- Recursos económicos de la institución.
- Habilidades tecnológicas de los funcionarios.
- Principales fuentes de datos.
- Objetivos de la unidad de información.
- Misión y visión de la unidad de información.

#### 3.1 Metodología

Evaluando las metodologías consideradas se puede determinar que todas pueden ser utilizadas para cumplir con los objetivos de esta investigación con un distinto grado de completitud y complejidad. Al realizar un análisis de las metodologías KDD y Semma se llegó a la determinación de que estas eran equivalentes entre sí: *Prueba* puede ser identificada con *Selección*, *Explorar* puede ser identificada con *Pre-procesamiento*, *Modificar* se identifica con *Transformación*, *Modelo* con *Minería de Datos*, *Evaluar* con *Interpretación/Evaluación*.

Sin embargo, por el nivel de detalle con el que describe las tareas en cada fase del proceso, que facilita su utilización a personas con menos experiencia y que busquen resultados óptimos (menos complejidad), porque incorpora actividades para la gestión del proyecto como gestión del tiempo, costo, riesgo (más completitud), se optó por utilizar la metodología CRISP-DM para la creación de una propuesta de implementación de Bibliomining en Bibliotecas escolares CRA en Chile.

## **3.2 Tecnologías asociadas**

### **3.2.1 Programas**

Como se mencionó anteriormente, Rapid Miner es un programa que funciona muy bien para pequeños grupos, en este caso una biblioteca escolar CRA. Es sencillo en la carga de datos gracias a sus operadores. No requiere del aprendizaje de ningún lenguaje de programación, por lo que no se necesitan mayores conocimientos de estadística ni informática para operarlo. Gracias a su versión gratuita, no requiere de la inversión de recursos económicos para su adquisición. Estos dos últimos puntos mencionados son muy importantes a considerar cuando se busca un programa para poner en práctica bibliomining en una biblioteca escolar CRA, ya que como se estableció en los criterios, se comienza de la base de que la institución cuenta con recursos limitados. Sus funcionarios, quienes serán los encargados de implementar este nuevo sistema, no poseen mayores conocimientos de estadística o informática, por lo que Rapid Miner es una solución amigable, práctica y gratuita pero de excelentes resultados.

### **3.2.2 Lenguajes**

Tal y como se mencionó en el punto anterior, dadas las características de Rapid Miner, no sería necesario en esta oportunidad contar con un lenguaje de programación como R o Python. Sin embargo, en el caso de que la biblioteca escolar CRA cuente con profesionales entendidos en materia de programación (informática) sería una gran ventaja evaluar la posibilidad de utilizar un programa que sí utilice lenguajes de programación como R o Python, dada la potencia de estos lenguajes. Ahora, si este no es el caso, se pone en marcha la implementación con Rapid Miner sin apoyarse de un lenguaje, lo que no quita solidez ni efectividad a la implementación.

## **3.3 Propuesta de implementación**

En esta etapa de la investigación, se realizará la tarea práctica de aplicar cada una de las fases de la metodología CRISP-DM al problema práctico que nos planteamos, que es la implementación de una propuesta de Bibliomining en bibliotecas escolares CRA. Se irán enumerando cada una de las fases de la metodología tal y como están numeradas en el punto 2.2 CRISP-DM.

### **1. Comprensión del negocio**

La finalidad de esta fase es determinar los objetivos y requisitos del proyecto desde una perspectiva de negocio, para más adelante poder convertirlos en objetivos desde el punto de vista técnico y finalmente en un plan de proyecto.

### **1.1 Determinar los objetivos del negocio**

El objetivo de esta metodología es obtener información fiable sobre los procesos realizados en una biblioteca escolar CRA, que permita mejorar los procesos administrativos realizados dentro de la institución.<sup>53</sup> Como primera tarea se recomienda estudiar la visión y la misión de la institución, encontrando aquí objetivos claros que permitan enfocar los intereses de la institución, como el tipo de alumno que se quiere formar, si se pretende realizar una formación humanista, científica, de un área técnica y/o con instrucción religiosa. Además del estudio de la planificación estratégica junto con el organigrama. Estas indicaciones son fundamentales para determinar los objetivos del negocio de la forma más concreta posible.

Una actividad que se recomienda es la realización de un FODA, herramienta de estudio de la situación de una empresa, institución, proyecto o persona, analizando sus características internas (Debilidades y Fortalezas) y su situación externa (Amenazas y Oportunidades) en una matriz cuadrada. A modo de ejemplo, una fortaleza sería contar con una colección organizada de manera prolija en una base de datos, que puede resultar en un rápido tratamiento de los datos. Una debilidad, siguiendo la misma línea, sería no contar con una base de datos, teniendo que realizar el proceso con información análoga, lo que no resulta excluyente del proceso, solo agrega tareas adicionales para compensar.

En cuanto a las características externas, una oportunidad podría ser la posibilidad de postular a un proyecto concursable de innovación en centros escolares, bibliotecas, etc. Para conseguir fondos adicionales. Como amenaza puede considerarse la falta de recursos para contratar personal de apoyo para la realización del proyecto.

---

<sup>53</sup> [https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC\\_Victor\\_Galan\\_Cortina.pdf](https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf)



Explicación gráfica del FODA<sup>54</sup>.

El siguiente paso es escoger a las personas que participarán en la aplicación de este proyecto, necesitando de al menos una persona que tenga una amplia trayectoria dentro de la institución, de forma que aporte con su experiencia para la toma de decisiones. También será fundamental contar con algún administrativo que realice labores en la biblioteca, preferentemente la persona encargada, así como la persona a cargo del presupuesto de la institución, para que formen parte del grupo de trabajo. Además de incluir un individuo que posea conocimientos intermedios en base de datos, como un informático, para que realice la tarea de extraer y tratar los datos, así como detectar los resultados finales.

<sup>54</sup> TIEMPO DE NEGOCIOS. (2017). *El análisis FODA para auditar una organización y su entorno*. Disponible en: <https://tiempodenegocios.com/analisis-foda/>

El análisis final lo ha de realizar el grupo de trabajo en su totalidad, siendo esta la manera más efectiva de llegar a resultados más cercanos a la realidad, ya que la visión de trabajadores expertos en distintas materias serán la clave para aplicar de forma práctica y realizable las soluciones posteriores.

Luego de formar el equipo de trabajo, es necesario seleccionar las fuentes de donde se extraerán los datos. Entre las principales fuentes de datos de una biblioteca escolar CRA se encuentran:

Libros de registros, Sistema integrado de gestión de bibliotecas (Opac, colecciones, ordenamiento de estanterías, copias, préstamos, usuarios, administrativos), metadata bibliográfica, información de adquisiciones (costos, fechas, proveedores, materiales adquiridos), datos demográficos (comuna, colegio, tipo de colegio, tipo de enseñanza, promedio simce).

### **Objetivos del negocio (en función de los desafíos)**

- Predecir el tiempo promedio de vida útil del material bibliográfico.
- Filtrar información sobre los procesos de compra y venta de adquisiciones, así como la relación precio/impacto de los materiales bibliográficos.
- Descubrir los patrones de comportamiento de los usuarios, distinguiendo el uso de recursos físicos e informáticos para realizar un servicio más personalizado.
- Evaluar el cumplimiento de los estándares establecidos por el MINEDUC por parte de las bibliotecas escolares CRA.

### **Criterios de éxito del negocio**

La aplicación se considera como exitosa sí:

- Se predice el tiempo de vida útil del material bibliográfico.
- Los materiales comprados cuentan con más préstamos.
- Se distinguen las necesidades de uso de los usuarios, aumentando las visitas del sitio y/o la biblioteca.
- Se cumple con los estándares del MINEDUC para bibliotecas escolares CRA.

## **2. Comprensión de los datos:**

Una vez identificados los datos de la fase previa, se procederá a la recopilación de estos mismos datos para su posterior análisis. En este punto, los datos alcanzados se describirán, así como se indagará en busca de nuevos datos y se verificará la calidad y relevancia de los mismos. Y una vez analizados y

verificados los datos recolectados, se comienza con la fase siguiente de acuerdo al modelo de CRISP-DM.

Como se hizo mención en el punto anterior, entre los datos a recopilar, se destacan la misión y visión de la Biblioteca escolar CRA, así como el organigrama de la institución, el tipo de usuario objetivo y la planificación estratégica entre otras.

Otro de los elementos a tomar en cuenta en esta etapa, es que, como su nombre lo indica, los datos deben ser comprendidos, esto implica que se conozcan las tablas de conversiones de los datos codificados, por ejemplo, o bien interpretados los datos que podrían causar confusión en los lectores de los mismos.

Para los efectos de esta propuesta, habrá que hacer mención también a que la recolección de los datos será tanto desde las fuentes de datos internas como las externas, de modo en que se pueda abarcar un alcance de un radio de interés de la biblioteca escolar CRA, implicando satisfacer necesidades de usuarios reales y/o potenciales. Por lo mismo, otro de los datos a los que se puede hacer mención es el lugar geográfico de la institución y los ya mencionados potenciales usuarios con sus respectivos perfiles.

### **3. Preparación de los datos:**

Como su nombre lo indica, ésta es la fase en la que los datos serán dispuestos para su posterior modelación. Como se mencionó en las fases de CRISP-DM del marco teórico (capítulo 2), en la etapa de preparación se busca construir el dataset. Es importante que la tarea de preparación de datos sea minuciosa y se lleve a cabo cuantas veces se considere necesario con el objetivo de transformar los datos iniciales que se recolectaron (tanto datos internos como externos) en datos limpios listos para las herramientas de modelado. De la misma manera, se requiere tanto de un material adecuado como de un personal capacitado para manejar el proceso.

Esto hace referencia a los dispositivos con los que se llevarán a cabo las bases de datos en los puntos siguientes y los profesionales o técnicos a cargo de implementarlas.

Dado todas las que se tienen que realizar, esta suele ser la etapa que toma más tiempo.

La preparación de datos cuenta con 5 etapas que se sugiere se realicen todas.

1. Lo primero es identificar los datos, ver dónde se encuentran. Suele suceder que los datos que se utilizarán están repartidos en planillas excel, sistemas privados de la institución, entre otros.
2. Una vez reconocidos los datos, se pasan a consolidar.
3. La tercera fase consiste en la edición de los datos para luego normalizarlos. Por ejemplo, al venir de distintas fuentes, puede que un rut se encuentre con puntos y guión en una planilla, sin embargo en el sistema privado de la institución no tendrá ni puntos ni guiones. Lo mismo pasa con fechas de nacimiento, algunas estarán dispuestas con el año, luego mes y finalmente día y en otra planilla podrían estar como día, mes, año y es fundamental que exista una normalización.
4. Es importante realizar una acuciosa revisión de datos faltantes y valores fuera de rango. Se debe establecer qué datos son fundamentales para la investigación y cuáles no. En caso de faltar datos que no serán significativos, se procede a eliminar al estudiante para no alterar la muestra. Sin embargo, si los datos faltantes son relevantes para la muestra, se completa con alguna de las alternativas que ofrece Rapid Miner, por ejemplo, si falta la edad de un alumno se saca un promedio de las edades que se tiene y se rellena el dato faltante.
5. Cuando se tienen todos los datos se pasa a la etapa final de limpieza de los mismos, para eliminar duplicados y asegurarse de que los datos están listos para pasar a la modelación.

Se sugiere utilizar una planilla excel para la preparación y consolidación de los datos. Esta misma planilla será la que se visualizará en la fase siguiente.

#### **4. Modelado:**

La fase de modelado consiste en definir para qué se utilizarán los datos, para crear patrones de comportamiento, generar reportes, entre otras cosas. Gracias a la fase 3, la preparación de los datos, en la etapa de modelado se trabaja con el set de datos ya limpio y editado.

- Lo primero que se sugiere hacer es definir claramente cuál será el fin último de estos datos, patrones de comportamiento, generación de reportes, pues en base a eso se escogerá la técnica de MD más adecuada para aplicar. Al existir varias técnicas de MD es importante no saltarse el punto de definición de la actividad para usar la técnica correcta.

- Luego de realizar el punto anterior, se crean 4 o 5 modelos con distintas técnicas de MD para ver cuál se ajusta bien y arroja mejores resultados.
- Finalmente, después de haber analizado qué modelo y técnica resulta más útil para la institución y el objetivo, se define dicho modelo y técnica para pasar a la evaluación.

## 5. Evaluación:

### Evaluación de los resultados

En esta etapa se analizarán los resultados obtenidos en la etapa anterior, con el objetivo de determinar si estos cumple con los “criterios de éxito” establecidos en la etapa 1, de esta forma se puede evaluar si el proceso de MD ha sido o no exitoso.

Una vez concluida la fase de modelado, la comprensión del negocio es determinante a la hora de decidir la forma de aplicar los resultados. Por esto se requiere una clara comprensión de los objetivos de la organización en que se ha implementado, por lo que también es importante visitar los objetivos del negocio antes identificados. Además de esto, es necesario tener identificado cuáles serán las variables que tendrán más peso según lo que se está analizando. En esta etapa ya se debería saber cuál herramienta es la mejor según el set de datos a partir de lo que se quiere realizar (establecer patrones o predecir comportamientos).

Para iniciar este proceso se puede apoyar respondiendo las siguientes preguntas:

- ¿Los resultados se expresan con claridad y de forma que se puedan presentar con facilidad?
- ¿Ha realizado descubrimientos especiales o particularmente relevantes que deba resaltar?
- ¿Puede evaluar los modelos y descubrimientos en función de su capacidad de poderse aplicar a los objetivos de la unidad? En general, ¿en qué medida estos resultados se adaptan a los objetivos de su unidad?
- ¿Qué cuestiones adicionales generan los resultados? ¿Cómo puede formular estas cuestiones en términos que favorezcan a la unidad?

### Proceso de revisión

El proceso de evaluación se complementa con la revisión del trabajo hecho para realizar un proceso de mejora de las acciones realizadas. En primer lugar

se deben resumir las actividades y decisiones realizadas en cada fase, incluyendo pasos para la preparación de datos, generación de modelos, etc.

Además, en cada fase, es necesario considerar si se sigue la línea de los objetivos previamente establecidos y realizar sugerencias para la mejora. Una vez se hayan evaluado los resultados, se debe realizar una lista con los modelos aprobados para incluirlos en el informe final. Los modelos escogidos son los que cumplen los requisitos de minería de datos y los objetivos de la organización.

Hecho esto es vital realizar las siguientes preguntas :

- ¿Esta fase ha contribuido al valor de sus resultados finales?
- ¿Existen formas de simplificar o mejorar esta fase u operación particular?
- ¿Cuáles fueron los fallos o errores cometidos en esa fase? ¿Cómo se pueden evitar la próxima vez?
- ¿Hay callejones sin salida, como modelos específicos que no ofrecen ningún resultado? ¿Existen formas de predecir esos callejones sin salida de forma que los esfuerzos se puedan dirigir con más productividad?
- ¿Se han producido sorpresas (buenas y malas) en esta fase? A posteriori, ¿existe alguna forma de predecir esas instancias?
- ¿Existen decisiones alternativas o estrategias que se puedan utilizar en una fase concreta? Registre esas alternativas para utilizarlas en proyectos de minería de datos futuros.

Para completar esta tarea de la forma más eficiente se recomienda tener en cuenta las siguientes recomendaciones:

- Cuando se produce un suceso inesperado en una fase diferente al CRISP-DM, hay que volver siempre al proceso de exploración.
- La preparación de los datos, especialmente de registros Web, requiere paciencia, ya que puede llevar mucho tiempo.
- Es esencial mantenerse centrado en el problema de negocio, porque una vez que los datos están preparados para el análisis, es muy fácil iniciar la construcción de modelos sin depender de la imagen mayor.

**6. Implementación:** Es en esta etapa donde los nuevos conocimientos se enfocan en implementar las mejoras a la organización. Una vez el modelo fue construido, validado y evaluado, se procede a pasmar todo lo aprendido en el proceso final. Hay que destacar también que este proceso, al tratarse de una etapa a largo plazo, deberá poder asentarse

con una buena mirada tanto de los implementadores como de sus beneficiarios.

En síntesis, en la fase de implementación, todas las preguntas desarrolladas en el punto anterior, una vez respuestas, comienzan a darle forma a esta nueva etapa.

Es ideal que los resultados, modelos y descubrimientos queden bien resueltos y, si es posible, resumidos, de modo que la implementación pueda aplicarse de manera más fácil y de acuerdo a los procesos anteriores que ya se han seguido. También será bien evaluada la creación de un plan de difusión para cada descubrimiento nuevo realizado, de modo que los estrategias de la organización los contemplen, para asegurar la implementación, en este caso, el de las bibliotecas escolares CRA, es necesario que el bibliotecario a cargo pueda revisar y dialogar los planes de difusión, para que sea partícipe de estas nuevas implementaciones.

Por otro lado, deben albergarse planes de difusión alternativos a los ya planteados, para que los tipos de resultados obtenidos en las fases anteriores tengan un mayor rango de cobertura. En caso de las bibliotecas escolares CRA, si se tratan los temas de misión y visión de la organización, por ejemplo, será necesario disponer de planes variables que posibiliten la continuidad de la investigación. El mismo fenómeno es el esperado cuando los datos recolectados tratan la colección de la biblioteca, el organigrama, la cantidad del personal, entre otras respuestas variables.

Para todas estas labores, no está demás decir que harán falta encargados de área, que bien puede ser un bibliotecario, para que supervise todas las aristas, debido a que se trata de una biblioteca escolar CRA, de modo que supervise el trabajo de los programadores o el personal que implemente todo lo anteriormente mencionado.

Finalmente, tener claro que los proyectos de minería de datos no finiquitan con la implementación de los modelos o de sus mejoras, pues deben ser documentados al tiempo que sus resultados deben ser presentados de manera adecuada para el usuario, haciendo aún mayor la retribución y empatía entre organización, personal y beneficiarios, al tiempo que se desempeñarán como frutos de proyección y comparación a futuro.

## Plan del proyecto<sup>55</sup>

Fase	Tiempo	Recursos	Riesgos
<b>Comprensión del negocio</b>	1 semana	Bibliotecarios, estudiantes, administrativos de la institución, tiempo.	*Objetivos demasiado ambiciosos. *Objetivos inconsecuentes con los intereses de la unidad.
<b>Comprensión de los datos</b>	3 semanas	Bibliotecarios, informáticos, fuentes de datos (en todos los formatos).	*Datos incompletos, problemas tecnológicos.
<b>Preparación de los datos</b>	5 semanas	Bibliotecario, computadores adecuados, conocimientos sobre bases de datos, asesor de minería de datos (si requiere).	*Problemas de datos. *Problemas tecnológicos.
<b>Modelado</b>	2 semanas	Bibliotecario, computadores adecuados, conocimientos sobre bases de datos, asesor de minería de datos (si requiere).	*Problemas tecnológicos. *Mal aplicación del modelo seleccionado.
<b>Evaluación</b>	1 semana	Bibliotecarios, estudiantes, administrativos de la institución, tiempo.	*Cambio en la inversión económica. *Incapacidad para implementar resultados. *No se cumplió con los objetivos establecidos. *Necesidad de empezar de nuevo con la metodología para cumplir con los objetivos.

<b>Implementación</b>	1 semana	Bibliotecarios, administrativos de la institución, asesor de minería de datos, tiempo de análisis de base de datos, recursos económicos.	*Cambio económico. *Incapacidad para implementar resultados.
-----------------------	----------	--	---

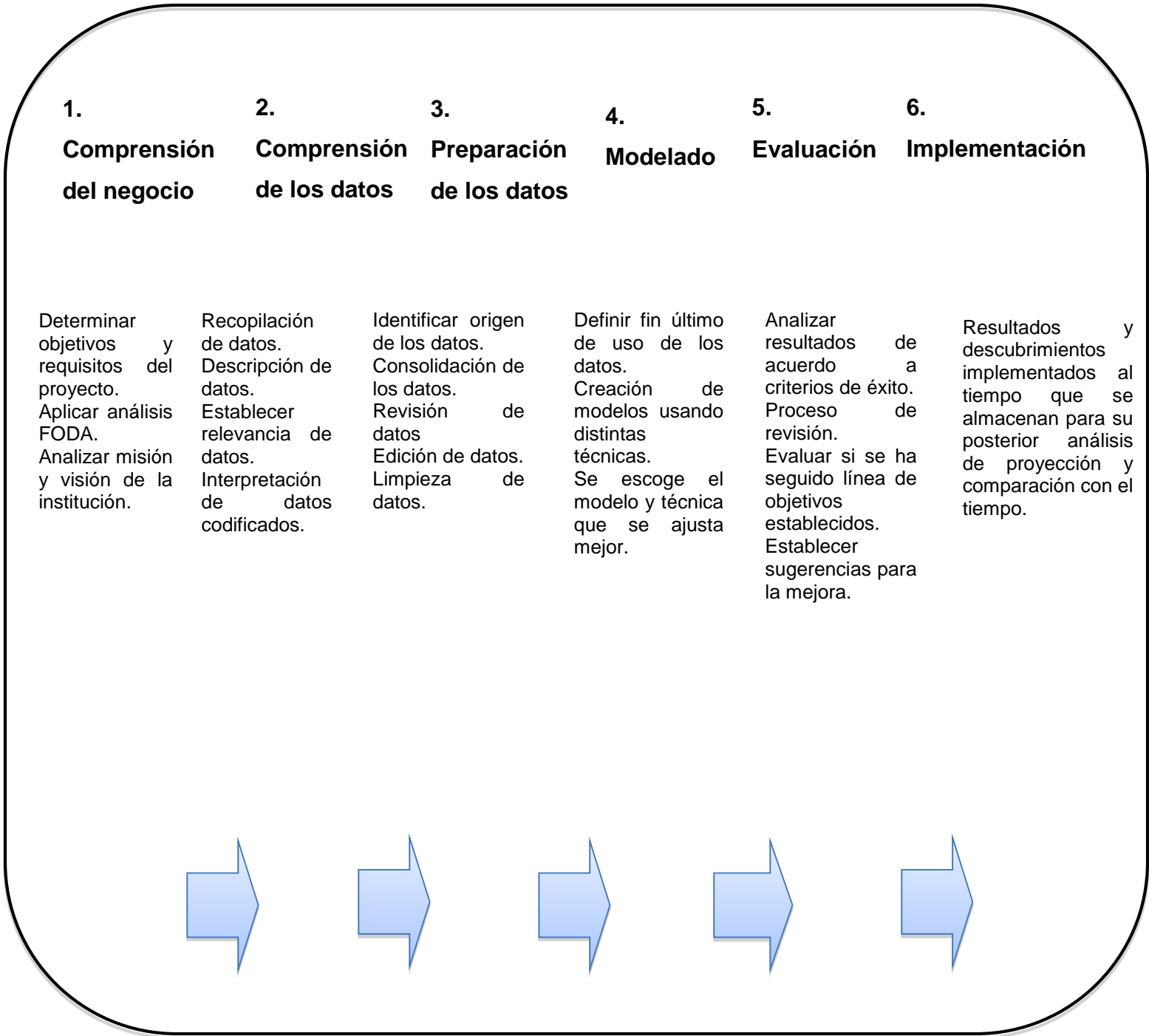


Fig. 13 Cuadro resumen de las fases de CRISP-DM aplicadas en bibliotecas escolares CRA.

## **Conclusiones**

La minería de datos tiene una indudable utilidad no sólo para empresas y temas de marketing. Su transversalidad permite aplicarla en colegios, universidades, centros de documentación, bibliotecas, entre otros muchos lugares de diversas características.

La MD vino a dar respuesta a un importante conflicto en materia de datos. Demostró que siempre habrá información escondida en una planilla. Existe mucha información oculta muy valiosa, que si es bien recuperada y utilizada, permite predecir eventos futuros, tomar mejores decisiones y ahorrar finanzas que podrían ser invertidas en otras cosas.

En esta oportunidad se analizó el término Bibliomining, es decir, la MD aplicada al mundo de las bibliotecas escolares CRA.

La MD no es un fin en sí misma, sino una herramienta que arroja una serie de resultados para dilucidar la realidad de una institución, permitiendo mejorar dicha realidad, la gestión y la toma de decisiones en el organismo.

Para finalizar se enfatiza que esta investigación es solo exploratoria. Su finalidad es establecer una metodología de implementación de Bibliomining, dejando los cimientos para que en una futura investigación se logre ir más allá y de ser posible se lleve a terreno lo que aquí se postula sólo como una idea.

## **Sugerencias**

Al ser Bibliomining un tema poco explorado en nuestra área, hay muchos aspectos por mejorar. El rol del bibliotecólogo ha ido mutando con el paso de los años y el avance de las tecnologías. Es, en este contexto, en que se hace indispensable que la educación del profesional de información se enmarque en el área de la estadística y la informática. Si las universidades que imparten la carrera dictaran cátedras obligatorias o electivas orientadas a la enseñanza de lenguajes de programación como Python o R, herramientas de predicción como Sas o Rapid Miner, el futuro encargado de la biblioteca CRA contará con mayores conocimientos y habilidades para desempeñar de mejor tarea su trabajo, hacerlo de manera integral y así brindar un mejor servicio a sus usuarios.

El constante perfeccionamiento de un profesional es algo fundamental, es por ello que se sugiere la puesta a disposición para la comunidad bibliotecaria más diplomados orientado a Bibliomining.

- La unificación de criterios en cuanto a la toma de decisiones previas a la realización de Bibliomining para que no se quede en algo de la institución y busque revolucionar el mundo bibliotecario.
- Las bibliotecas escolares CRA deberían automatizar los sistemas de ingreso de material, préstamo y devolución al interior de sus unidades, para que la aplicación de bibliomining sea más fácil y efectiva.

## Bibliografía

- Albashrawi, M. (2016). *Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015*. Journal of Data Science, 14(3), 553-569.
- Angúlo, F., & Sergio, E. (2012). *Modelo para la automatización del proceso de determinación de riesgo de deserción en alumnos universitarios*.
- Beltrán, David. Poveda, Diego.(2010). *Rapid Miner*.
- Brunner, José Joaquín (2008). *Problemas, realidades y desafíos de la Educación en Chile*.
- Candás Romero, Jorge (2006). *Minería de datos en bibliotecas : bibliominería*. BiD: textos universitarios de biblioteconomía i documentació, núm. 17 (diciembre) . <<http://bid.ub.edu/17canda2.htm>>. [Consulta: 25-04-2018].
- Castillo, J. J., & Cardenas, M. E. (2017, August). *Aplicaciones de análisis de información textual: corpus lingüísticos*. In XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires).
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.
- Cía, Juan F. (2015). *El Ranking de las mejores soluciones de análisis predictivo para empresas*. Disponible en: <<https://bbvaopen4u.com/es/actualidad/el-ranking-de-las-mejores-soluciones-de-analisis-predictivo-para-empresas>>
- Cobos, C., Zuñiga, J., Guarín, J., León, E., & Mendoza, M. (2010). *CMIN-herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos*. Ingeniería e Investigación, 30(3).
- De-la-Puente, M. (2010). *Bibliominería: bibliometría y minería de datos*.
- Díaz, María Victoria. (2016). *Metodología CRISP-DM – Parte 1*. Disponible en: <<http://smartbasegroup.com/metodologia-crisp-dm-parte-i/>>.
- Elaboración propia basado en: G2 CROWD. *Compare RapidMiner vs SAS*. Disponible en: <<https://www.g2crowd.com/compare/rapidminer-studio-vs-sas-business-intelligence>>
- Elaboración propia. Basado en: Beltrán, David y Poveda, Diego. (2010). *Rapid Miner*.
- G2 CROWD. (2018). *Compare RapidMiner vs SAS*. Disponible en: <<https://www.g2crowd.com/compare/rapidminer-studio-vs-sas-business-intelligence>>.
- Gondar Nores, J.-E.(2004). *Metodologías para la Realización de Proyectos de Data Mining*. Disponible en: <<http://www.estadistico.com/arts.html?20040426>>.
- Gorbea-Portal, S. (2013). *Tendencias transdisciplinarias en los estudios métricos de la información y su relación con la gestión de la información y del conocimiento*. Perspectivas em Gestão & Conhecimento, 3(1), 13-27.
- Gutiérrez, J. A., & Molina, B. (2016). *Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales*. Revista Ontare, 3(2), 33-51.

- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier. La traducción es nuestra.
- Haro, Pérez, Saquicela. (2016). *Bibliomining para descubrir reglas de asociación en el Centro de Documentación Regional "Juan Bautista Vázquez*.
- Jans, M., Van Der Werf, J. M., Lybaert, N., & Vanhoof, K. (2011). *A business process mining application for internal transaction fraud mitigation*. *Expert Systems with Applications*, 38(10), 13351-13359.
- Koenig, M. E. D. (1985). *Bibliographic Information Retrieval Systems and Database Management Systems*. *Information Technology and Libraries*, 4, 247-272.
- López García, D. (2013). *Análisis de las posibilidades de uso de Big Data en las organizaciones* (Master's thesis). Disponible en: <https://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADa.pdf?sequence=1>.
- Martinez, G. R. S., & Ocampo, C. A. (2009). *Técnicas de clasificación y análisis de representación del conocimiento para problemas de diagnóstico*. *Scientia et Technica*, 2(42).
- MINEDUC. (2011). *Estándares para las Bibliotecas Escolares CRA*.
- MINEDUC. *Bibliotecas Escolares CRA*.
- Ministerio de Educación, (2014). *20 años de bibliotecas escolares CRA*. Disponible en: [https://issuu.com/bibliotecas\\_cra/docs/20\\_anos\\_de\\_bibliotecas\\_escolares\\_cr\\_a14a10b868c559](https://issuu.com/bibliotecas_cra/docs/20_anos_de_bibliotecas_escolares_cr_a14a10b868c559).
- Ministerio de Educación. *Aplicación de encuesta censal en las bibliotecas escolares cra de los establecimientos educacionales subvencionados del país*. Disponible en: [http://www.bibliotecas-cra.cl/sites/default/files/publicaciones/estudioencuesta\\_censal\\_2011.pdf](http://www.bibliotecas-cra.cl/sites/default/files/publicaciones/estudioencuesta_censal_2011.pdf)
- Ministerio de Educación. *Misión*. Disponible en: <http://www.bibliotecas-cra.cl/?q=mision>.
- Ministerio de Educación. *Software Abies 2.0*. Disponible en: <http://www.bibliotecas-cra.cl/abies2>.
- Nicholson, S. (2003). *Bibliomining for automated collection development in a digital library setting: Using data mining to discover Web-based scholarly research works*. *Journal of the Association for Information Science and Technology*, 54(12), 1081-1090.
- Nicholson, Scott. (2004). *The Bibliomining Process: Data Warehouse and Data Mining for Library Decision Making*.
- Nicholson, Scott. Stanton, Jeffrey. (2003). *Gaining Strategic Advantage Through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries*. Disponible en: <https://surface.syr.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1112&context=istpub>

Papatheodorou, Christos; Kapidakis, Sarantos; Sfakakis, Michalis; Vassilou, Alexandra (2003). *Mining user communities in digital libraries*. Information technology and libraries, December, vol. 22, no. 4, p. 152-157.

Paradis, Emmanuel. (2013). *R para Principiantes*.

Pérez Kauscuhs, Gabriela (2015). *El rol de las bibliotecas escolares en el sistema educativo chileno: descripción, desafíos y recomendaciones para su mayor efectividad en el incremento de oportunidades educativas*.

Pérez, W., Haro, V., & Saquicela, V. (2017). *Bibliomining: Aplicación de text mining para descubrir preferencias de usuarios en el Centro de Documentación Regional "Juan Bautista Vázquez"*. *Maskana*, 8, 135-144.

Python Software Foundation. Disponible en: <<https://www.python.org/about/>>

Quiroz Gil, N. L., & Valencia, C. A. (2012). *Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim*. *Revista Interamericana de Bibliotecología*, 35(1).

R Development Core Team. (2000). *Introducción a R*.

RapidMiner Community. (2018). Disponible en: <<https://community.rapidminer.com/>>

SAS Enterprise Miner. (2006). *SEMMA*. Disponible en: <[http://faculty.smu.edu/tfomby/eco5385\\_eco6380/data/SPSS/SAS%20%20SEMMA.pdf](http://faculty.smu.edu/tfomby/eco5385_eco6380/data/SPSS/SAS%20%20SEMMA.pdf)>.

Sosa, M. O., Chesñevar, C. I., & Sosa Bruchmann, E. C. (2016). *Framework para data mining educativo: formalización y aplicaciones*.

Virseda, F. & Ramón, J. (2011). *Minería de Datos y Aplicaciones*. Disponible en: <<http://www.it.uc3m.es/jvillena/irc/practicas/06-07/22.pdf>>.

YOUR FREE TEMPLATES. (2017). *Phases of SEMMA*. Disponible en: <https://yourfreetemplates.com/free-data-mining-template/>.

## **Glosario**

**Bibliomining:** Se entenderá por bibliomining como la minería de datos aplicada al campo de las bibliotecas.

**Big data:** Gran volumen de datos difíciles de procesar que crecen de manera exponencial.

**CMIN:** Herramienta basada en CRISP-DM que orienta el desarrollo de proyectos.

**CRA:** Centro de Recursos del Aprendizaje.

**CRISP DM:** Modelo utilizado para MD.

**Data Warehouse:** Gran fuente de datos que integra datos desde otras fuentes de datos que depura la información recolectada.

**Jupyter:** Entorno de programación para Python.

**KDD:** Metodología de MD que obtiene conocimiento gracias a datos rescatados de un Data Warehouse.

**Minería de datos:** Técnica de extracción de datos para su posterior interpretación con el fin de predecir patrones y comportamientos.

**OPAC:** Online Public Acces Catalog. Catálogo en línea.

**Python:** Lenguaje de programación que permite integrar sistemas.

**R:** Lenguaje de programación.

**Rapid Miner:** Programa comercial con versión gratuita que permite realizar MD. Es un programa intuitivo.

**R Studio:** Entorno de programación para visualizar e interpretar R.

**SAS:** Programa comercial para el desarrollo de MD.

**SEMMA:** Herramienta desarrollada para aplicar variadas metodologías para la MD.