

### CAPITULO III. CRISP-DM: CONOCIMIENTO Y COMUNICACIÓN DE UNA METODOLOGÍA PARA MINERÍA DE DATOS.

**Autor:**

**Marcos Antonio Espinoza Mina, Mgs.**  
Docente Facultad de Sistemas y Telecomunicaciones  
Universidad Tecnológica ECOTEC  
Universidad Agraria del Ecuador  
[mespinoza@ecotec.edu.ec](mailto:mespinoza@ecotec.edu.ec)

#### INTRODUCCIÓN

La información se ha convertido en uno de los recursos más valiosos de las organizaciones. Con el pasar del tiempo el volumen de datos crece en diferentes ubicaciones de la empresa y en muchos casos quedan aislados, siendo muy necesaria la unificación y la consolidación que permita la correcta toma de decisiones.

Muchas organizaciones aún enfrentan dificultades para determinar cómo administrar su información empresarial y aprovechar su uso para impulsar sus decisiones (Calzada & Abreu, 2009). Para obtener conocimiento es necesario partir de la materia prima, que son los datos, los cuales se encuentran disponibles en gran cantidad gracias a las tecnologías de información y las comunicaciones. Estos datos por lo general se encuentran en forma no refinada y para poder analizarlos con fiabilidad es necesario que exista una cierta estructuración y coherencia entre los mismos (Romero & Paredes, 2017). Para subsanar estos problemas se presentan conceptos como BI, DM y descubrimiento de conocimiento en bases de datos (KDD, del inglés Knowledge Discovery in Databases).

La información es una base muy importante para que la organización pueda detectar oportunidades y analizar la factibilidad y rentabilidad de llevarla a cabo las mejoras a través de un proyecto. El éxito de los proyectos depende de la sincronización que se pueda tener entre diferentes factores como los humanos,

los técnicos, los económicos, tiempo, alcance, entre otros (Gallego & Hernández, 2015). Dentro de los recursos técnicos se destaca la información, volviéndose recursiva la necesidad de contar con ella y los medios para proveerla.

DM apoya a las organizaciones en diferentes áreas, por ejemplo, dentro del proceso de comercialización de un producto o servicio desde el fabricante o proveedor hasta el usuario o consumidor final. Al aplicar técnicas de agrupamiento de DM y análisis estadísticos para la obtención de información referente a los indicadores comerciales, es posible agilizar la definición veraz y oportuna de los indicadores para que sea más eficiente la toma de decisiones, construyendo una solución de DM para BI aplicado a canal detallista. La información relacionada con dichos indicadores pueda ser obtenida de manera eficiente veraz y oportuna. De esta forma, los usuarios podrán tomar decisiones que concuerden con la realidad del entorno comercial de la empresa (Flores, Pulido, & de la Rosa, 2016).

Para una empresa el tomar una decisión sin el conocimiento profundo de la información implica la posibilidad de errar en la toma de decisiones, y esta genera el costo de los valores invertidos para poner en marcha un plan que busque la fidelidad de los clientes o capturar nuevos, o cautivar a un nuevo nicho de mercado (Naranjo & Sierra, 2009).

Hoy en día las modernas herramientas de DM permiten extraer conocimiento de los datos con los que se cuenta, sin importar si están ordenados o no se encuentren fácilmente; ayudan a integrar las diversas fuentes de datos para que la alta gerencia pueda tomar información en tiempo real, desde cualquier lugar, desde cualquier dispositivo y dar a conocer a los involucrados la toma de decisiones hacia cualquiera de las áreas de la organización.

Las presiones económicas y la competencia han llevado a los directores de marketing a invertir en campañas dirigidas con una estricta y rigurosa selección de contactos. Estas campañas directas se pueden mejorar mediante el uso de técnicas de BI y DM (Moro, Laureano, & Cortez, 2011).

La capacidad para almacenar datos ha crecido en los últimos años a velocidades exponenciales. En el otro extremo, la capacidad para procesar esta enorme cantidad de datos para utilizarlos eficazmente no ha ido a la par. Por este motivo, DM se presenta como una tecnología de apoyo para explorar, analizar, comprender y aplicar el conocimiento obtenido de grandes volúmenes de datos. (Camana, 2016).

Las metodologías para el desarrollo e ingeniería de software han logrado estandarizar los procesos de software. DM no puede tratarse con estas mismas metodologías ya que tiene sus propias características. Existen metodologías dirigidas específicamente para ejecutar este tipo de proyectos, entre las más empleadas se encuentran CRISP-DM y SEMMA (Sample, Explore, Modify, Model, Assess).

Desde los primeros años de divulgación, CRISP-DM obtuvo apoyos de empresas privadas y organismos públicos, por su calidad y efectividad del método y hoy en día es adoptado por muchos organismos y empresas.

Otra metodología muy considerada es SEMMA, que fue desarrollada como un agregado al software estadístico SAS. CRISP-DM a pesar de ser parte del paquete de SPPS de IBM no está fuertemente vinculada con ella, cumpliendo con el propósito de ayuda a la labor de DM.

El presente trabajo recopila la definición, características e implicaciones de las herramientas de BI y de DM. Describe de forma resumida las fases de la metodología SEMMA y de forma un poco más ampliada, por su mayor importancia la de CRISP-DM; ambas como marcos de trabajo para proyectos de DM. Se expone a la comunidad profesional una metodología con suficiente nivel de madurez que permita la integración de DM adecuada, como parte de una solución de inteligencia de negocios, y de esta forma las empresas y organizaciones tengan a su alcance este tipo de solución y metodología, viabilizando el incremento de su productividad, lo que las llevará a ser más competitivas.

A continuación, se presenta a través de evaluaciones de diversos autores la información más relevante de BI, DM, KDD y dos de las más importantes metodologías para DM, con una búsqueda exhaustiva y cuidadosa de literatura científica y técnica actualizada de fuentes primarias y secundarias a nivel global. Se presenta descriptivamente a CRISP-DM, una de las metodologías más difundidas en la actualidad para proyectos de minería de datos, dejando abierta una puerta de discusión sobre el enfoque en el que debería ser considerada.

### **3.1 BI (inteligencia de negocios).**

Las plataformas de BI son un tipo de solución con diferentes aplicaciones, diseñadas para colaborar con el procesamiento de la información en los procesos de las organizaciones. Específicamente se trata de herramientas de asistencia para el análisis y la presentación de los datos (Rodríguez & Cortés, 2012). Se entiende por BI como el conjunto de aplicaciones, tecnologías, herramientas, métodos y estrategias, enfocadas a la toma de decisiones. Haciendo posible comparar, evaluar, analizar y responder a cuestionamientos, mediciones o indicadores empresariales.

Estudios indican que existe un incremento continuo en las publicaciones relacionadas con el concepto de BI en América Latina como conjunto y en cada país por separado, aunque existen algunos países que su crecimiento es lento. Los componentes de Inteligencia de Negocios que reciben mayor atención por parte de los investigadores son DM y las 'bodegas de datos', pero de igual forma la investigación se distribuye entre los otros diferentes conceptos que este término sombrilla encierra (Bustamante, Galvis, & Gómez, 2016).

### **DM (minería de datos) y KDD (descubrimiento de conocimiento en bases de datos).**

Dos objetivos clásicos de DM son el de sintetizar la información y descubrir un nuevo conocimiento (Alvarado & Cuervo, 2013). DM y el análisis predictivo satisfacen la demanda de herramientas para la toma de decisiones en tiempo real. Aunque DM ha sido un espacio de soluciones definido desde los años

noventa, recién en los últimos años se ha mejorado su proceso para crear patrones y tendencias en el entorno empresarial con base en la integración de datos de clientes, proveedores, empleados y productos.

Las técnicas utilizadas en DM están muy relacionadas con las técnicas estadísticas tradicionales, tales como la regresión de lineal o análisis de series en el tiempo, pero incorporadas y acompañadas de un conjunto mucho más rico y flexible de modelos.

KDD es básicamente un proceso automático en el que se combinan descubrimiento y análisis. KDD se ha desarrollado y continúa desarrollándose con base en las investigaciones realizadas en los campos del aprendizaje de máquina, reconocimiento de patrones, bases de datos, estadística, inteligencia artificial, sistemas expertos, visualización de datos y computación de alto rendimiento. La meta común es la extracción de conocimiento de los datos en el contexto de grandes bases de datos (Timarán, Hernández, Caicedo, Hidalgo, & Alvarado, 2016). KDD consiste en el proceso de extracción de información implícita poco común, desconocida, y potencialmente útil de los datos.

Se encuentra literatura que plantea el concepto de KDD similar al de DM, sin embargo algunos autores consideran que el primero es más amplio y abarcador que el segundo, donde DM sólo se refiere exclusivamente al conjunto de algoritmos y métodos empleados para extraer el conocimiento y forma parte del proceso del KDD. Se debe entender entonces que DM es la parte de extracción de conocimiento dentro del proceso global KDD.

Las metodologías de SEMMA y CRISP-DM pueden ser vistas como una implementación del proceso KDD (Azevedo & Filipe, 2008).

### **3.2 Metodologías**

Para la implementación de tecnologías en el área de informática es necesario contar con una metodología acorde al tipo de proyecto en el que se quiera trabajar. Los métodos aplicados deben ser seleccionados a partir de las

experiencias y tomando lo mejor de los procedimientos exitosos o más populares. Contar con metodologías dentro de los procesos es tan importante como la carta de presentación de la organización.

La utilización de una metodología estructurada y organizada presenta algunas ventajas para la realización de proyectos de DM, como es la facilidad en la realización de nuevos proyectos con características similares, sencillez en la planificación, dirección y seguimiento del proyecto (Rodríguez, Álvarez, Mesa, & González, 2003).

Al comparar los procesos de minería de datos KDD, CRISP-DM y SEMMA, se revela que la mayoría de los investigadores y expertos en DM siguen el modelo de proceso KDD porque es más completo y preciso. En contraste CRISP-DM y SEMMA son principalmente orientados a las empresas. SEMMA es utilizado por SAS y tiene una destacada participación, sin embargo, se demuestra que CRISP-DM es más completo que SEMMA. Estos modelos de procesos son una guía y ayuda a las personas y expertos a saber la forma en que pueden aplicar la minería de datos en escenarios prácticos (Shafique & Qaiser, 2014).

Se opta por el uso de la metodología de desarrollo a CRISP-DM por la independencia de esta metodología con respecto a las herramientas tecnológicas a utilizar en la explotación de datos; por ser de libre acceso, orientada al negocio y finalmente debido a ser la más completa de las metodologías evaluadas, ya que incluye además de los procesos de desarrollo, una fase preliminar dedicada al entendimiento del negocio que no es contemplada por el resto de las metodologías evaluadas (Vanrell, Bertone, & García, 2012).

Algunos modelos conocidos como metodologías son en realidad un modelo de proceso: un conjunto de actividades y tareas organizadas para llevar a cabo un trabajo. La diferencia fundamental entre metodología y modelo de proceso radica en que el modelo de proceso establece qué hacer, y la metodología especifica cómo hacerlo. Una metodología no solo define las fases de un proceso sino también las tareas que deberían realizarse y cómo llevar a cabo las mismas.

CRISP-DM es actualmente la guía de referencia más utilizada en el desarrollo de proyectos de minería de datos. Estructura el proceso en fases cuya sucesión no es necesariamente rígida. Cada fase se descompone en varias tareas generales de segundo nivel. CRISP-DM establece un conjunto de tareas y actividades para cada fase del proyecto pero no especifica cómo llevarlas a cabo, sin embargo podría ser considerada una metodología, por el nivel de detalle con el que describen las tareas en cada fase del proceso, y porque incorporan actividades para la gestión del proyecto como gestión del tiempo, costo, riesgo (Moine, Gordillo, & Haedo, 2011).

El modelo genérico de proceso CRISP-DM es útil para la planificación, comunicación dentro y fuera del equipo del proyecto y documentación. Las listas de verificación genéricas son útiles incluso para personas con experiencia. El modelo de proceso genérico proporciona una excelente base para el desarrollo de un modelo de proceso especializado que prescribe los pasos a seguir en detalle y que da consejos prácticos para todos estos pasos (Wirth & Hipp, 2000).

### 3.3 SEMMA

La metodología SEMMA (SAS Institute Inc, 2017) facilita la aplicación de técnicas estadísticas exploratorias y de visualización; permite la selección y transformación de las más significativas variables predictivas; modela las variables para predecir resultados y confirmar la exactitud de un modelo. Consta de seis fases: muestreo, exploración, manipulación, modelado y valoración; ver figura 1.

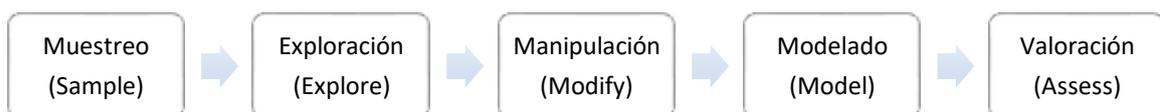


Figura 1. Fases de metodología SEMMA

SEMMA inicia con la extracción de la población muestral. La representatividad de la muestra es indispensable ya que de no cumplirse invalida todo el modelo y

los resultados dejan de ser admisibles. Regularmente se aplica el muestreo aleatorio simple. Se debe establecer para cada muestra el nivel de confianza de la muestra.

La segunda fase es la exploración de la información disponible con el fin de simplificar en lo posible el problema para optimizar la eficiencia del modelo. Para lograr este objetivo se propone la utilización de herramientas de visualización o de técnicas estadísticas que ayuden a poner de manifiesto relaciones entre variables, con esto se determina cuáles son las variables explicativas que van a servir como entradas al modelo.

La tercera fase de la metodología consiste en la manipulación de los datos, de forma que se definan y tengan el formato adecuado los datos que serán introducidos en el modelo.

Posteriormente se procede al análisis y modelado de los datos. Consiste en establecer una relación entre las variables explicativas y las variables objeto del estudio, que permitan inferir el valor de estas con un nivel de confianza determinado.

Las técnicas utilizadas para el modelado de los datos incluyen métodos estadísticos tradicionales como análisis discriminante, métodos de agrupamiento, y análisis de regresión, así como técnicas basadas en datos tales como redes neuronales, técnicas adaptativas, lógica difusa, árboles de decisión, reglas de asociación y computación evolutiva.

La última fase consiste en la valoración de los resultados mediante el análisis de bondad del modelo o modelos contrastados con otros métodos estadísticos o con nuevas poblaciones muestrales.

### **3.4 CRISP-DM**

Los orígenes de CRISP-DM, se remontan hacia finales de los noventa cuando un importante consorcio de la Unión Europea conformado por NCR (Dinamarca),

AG(Alemania), SPSS (Inglaterra), OHRA (Holanda), Teradata, y Daimler-Chrysler, propone a partir de diferentes versiones de KDD el desarrollo de una guía de referencia de libre distribución denominada CRISP-DM (Cross Industry Standard Process for Data Mining).

CRISP-DM es una metodología con propósitos generales para cualquier proyecto de MD; plantea ideas que deben parametrizarse para cada entorno de ejecución, quitando algunas cosas y adicionando otras, según sea la naturaleza y los objetivos de cada proyecto. Propone modelos genéricos que deben ser adaptados.

El ciclo de vida de CRISP-DM se ha convertido en el "estándar de oro" comúnmente aceptado de buenas prácticas en la metodología DM, a pesar de que proporciona poca o ninguna orientación o marco para el aspecto crucial en la elección del modelo DM. (Rennolls & Al-Shawabkeh, 2008).

La metodología CRISP-DM es capaz de estructurar el ciclo de vida de un proyecto en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo. Es de libre distribución y permite trabajar con cualquier herramienta para el desarrollo del proyecto (Montero & Plasencia, 2016).

El modelo de referencia CRISP-DM da una visión general del ciclo de vida de un proyecto de minería de datos, contiene las fases con sus objetivos, las tareas y las relaciones entre éstas, y las instrucciones paso a paso que se deben llevar a cabo (Cobos, Zuñiga, Guarín, León, & Mendoza, 2010). Las fases son seis en total: comprensión del negocio, comprensión de los datos, preparación de datos, modelado, evaluación y distribución, las cuales interactúan entre ellas de forma iterativa; ver figura 2.

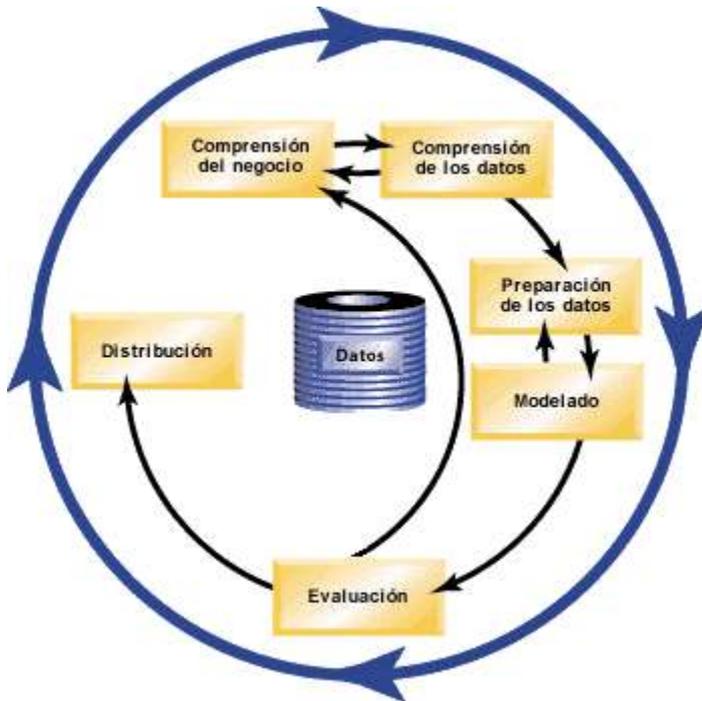


Figura 2. El ciclo vital de CRISP-DM.

Fuente: \*Tomado del manual CRISP-DM de IBM SPSS Modeler.

En CRISP-DM cada una de sus fases se encuentra claramente estructurada definiendo de tal forma las actividades y tareas que se requieren para lograr el objetivo planteado es decir es la más completa entre las metodologías comparadas, es flexible por ende se puede hacer uso de cualquier herramienta de minería de datos (Jaramillo & Arias, 2015). Cada fase tiene tareas generales y actividades específicas que generan un conjunto de resultados concretos, parte de estos detalles se reflejan de forma resumida en la Tabla 1.

Tabla 1. Fases y tareas de metodología CRISP-DM.

| Fase                           | Tareas                                     | Sub-tareas/Salidas  |
|--------------------------------|--|---|
| <b>Comprensión del negocio</b> | Determinación de los objetivos comerciales | <ul style="list-style-type: none"> <li>• Compilación de la información de la empresa</li> <li>• Definición de los objetivos comerciales</li> <li>• Criterios de rendimiento comercial</li> </ul>                                |
|                                | Valoración de la situación                 | <ul style="list-style-type: none"> <li>• Inventario de recursos</li> <li>• Requisitos, supuestos y restricciones</li> <li>• Riesgos y contingencias</li> <li>• Terminología</li> <li>• Análisis de costes/beneficios</li> </ul> |

## RETOS Y PERSPECTIVAS DE LAS TECNOLOGÍAS DE INFORMACIÓN

---

|                                 |  |  |
|---------------------------------|--|--|
| <b>Comprensión de los datos</b> | Determinación de los objetivos de minería de datos | <ul style="list-style-type: none"> <li>• Objetivos de minería de datos.</li> <li>• Criterios de rendimiento de minería de datos.</li> </ul>  |
|                                 | Producción de un plan de proyecto                  | <ul style="list-style-type: none"> <li>• Escritura del plan de proyecto</li> <li>• Plan de proyecto de muestra.</li> <li>• Valoración de herramientas y técnicas.</li> </ul>           |
|                                 | Recopilación de datos iniciales                    | <ul style="list-style-type: none"> <li>• Escritura de un informe de recopilación de datos.</li> </ul>  |
|                                 | Descripción de los datos                           | <ul style="list-style-type: none"> <li>• Escritura de un informe de descripción de datos.</li> </ul>   |
|                                 | Exploración de datos                               | <ul style="list-style-type: none"> <li>• Escritura de un informe de exploración de datos.</li> </ul>   |
| <b>Preparación de datos</b>     | Verificación de calidad de datos                   | <ul style="list-style-type: none"> <li>• Escritura de un informe de calidad de datos</li> </ul>  |
|                                 | Selección de datos                                 | <ul style="list-style-type: none"> <li>• Inclusión o exclusión de datos.</li> </ul>  |
|                                 | Limpieza de datos                                  | <ul style="list-style-type: none"> <li>• Escritura de un informe de limpieza de datos.</li> </ul>  |
|                                 | Construcción de nuevos datos                       | <ul style="list-style-type: none"> <li>• Derivación de atributos.</li> </ul>   |
|                                 | Integración de datos                               | <ul style="list-style-type: none"> <li>• Tareas de integración</li> </ul>  |
| <b>Modelado</b>                 | Formato de datos                                   | <ul style="list-style-type: none"> <li>• Aplicación de formato</li> </ul>  |
|                                 | Selección de técnicas de modelado                  | <ul style="list-style-type: none"> <li>• Selección de las técnicas de modelado correctas</li> <li>• Modelado de supuestos</li> </ul>   |
|                                 | Generación de un diseño de comprobación            | <ul style="list-style-type: none"> <li>• Escritura de un diseño de comprobación.</li> </ul>  |
| <b>Evaluación</b>               | Generación de los modelos                          | <ul style="list-style-type: none"> <li>• Configuración de parámetros</li> <li>• Ejecución de los modelos</li> <li>• Descripción de modelo</li> </ul>                                   |
|                                 | Evaluación del modelo                              | <ul style="list-style-type: none"> <li>• Evaluación global del modelo</li> </ul>   |
|                                 | Evaluación de los resultados                       | <ul style="list-style-type: none"> <li>• Constatación de que los resultados del proyecto cumplen con los criterios de rendimiento.</li> </ul>  |
|                                 | Proceso de revisión                                | <ul style="list-style-type: none"> <li>• Creación de resumen de actividades y decisiones de cada fase</li> <li>• Determinación de cuestionamientos y sugerencias de mejora.</li> </ul> |
|                                 | Determinación de los pasos siguientes              | <ul style="list-style-type: none"> <li>• Definición de posibles acciones y decisiones en términos de objetivos comerciales de minería de datos.</li> </ul>                             |

---

|                     |   |    |   |
|---------------------|---|----|---|
| <b>Distribución</b> | Planificación de distribución                 | de | • Creación de plan de distribución completo y preciso                                 |
|                     | Planificación del control y del mantenimiento |    | • Generación de estrategia para monitorear y mantener la solución                     |
|                     | Creación de un informe final                  |    | • Preparación de una presentación final   |
|                     | Revisión final del proyecto                   |    | • Determinación lo fructífero del proyecto y la obtención de los resultados esperados |

Fuente: Elaboración propia basado en el manual CRISP-DM de IBM SPSS Modeler.

### CONCLUSIONES

Las herramientas de BI proveen información o conocimiento nuevo y relevante que afecta a la productividad y competitividad de la organización.

Las bases de datos operativas son la materia prima para DM y obligatoriamente es a través de una metodología que se pueden construir adecuados modelos que permitan desarrollar diversas campañas ajustadas como las de promoción, marketing o fidelización.

Son algunas las técnicas de modelado que emergen y se encuentran disponibles para llevar a cabo DM; este incremento es muy importante. También es cierto que la base de todas estas herramientas, es decir, las técnicas de regresión, el análisis exploratorio de datos y aquellos sencillos gráficos estadísticos ayudan a revelar información importante y patrones ocultos; sin embargo, ante la alta competitividad y el crecimiento de los datos, ningún único método ni una sola herramienta, peor manual no automatizada, proveería la mejor solución; las organizaciones requieren obligatoriamente de un amplio conjunto de ellas y apoyarse en una metodología para su implementación.

Otro punto a considerar encontrado en la literatura revisada es que para que el software de BI se convierta en una verdadera solución, se requiere que las

organizaciones cuenten con una cultura y procesos adecuados, la infraestructura técnica bien configurada y las habilidades necesarias para una interpretación y toma de decisiones ágil y eficiente.

De acuerdo a lo revisado, CRISP-DM propone una lógica más amplia que SEMMA, busca el entendimiento del negocio y el por qué se debería implementar un proyecto de DM. Al ser publicada y distribuida libremente CRISP-DM en su sitio web, puede ser implementada por cualquier organización que tenga interés en desarrollar un proyecto DM; estos elementos justifican el por qué sigue siendo considerada por los especialistas como una propuesta adecuada para la implementación de soluciones de BI.

El presente trabajo motivará a los profesionales del medio informático a insistir a la alta gerencia en el inicio de proyectos DM dentro de las organizaciones, así como la aplicación de una metodología para su desarrollo; derivados de este documento se generarán posibles investigaciones que sigan confirmando o muestren un tentativo declive del uso de CRISP-DM como metodología para DM.

### REFERENCIAS

- Alvarado, J. A., & Cuervo, C. (2013). Extracción de Funciones de un Cargo usando Minería de Texto en Correos Electrónicos. *Información Tecnológica*, 24(5), 61-68. <https://doi.org/10.4067/S0718-07642013000500008>
- Azevedo, A., & Filipe, M. (2008). KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW. En *IADIS European Conference Data Mining 2008* (pp. 182-185).
- Bustamante, A., Galvis, E., & Gómez, L. C. (2016). Perfil de la investigación sobre inteligencia de negocios en América Latina. *Revista UIS Ingenierías*, 15(1), 41-51. <https://doi.org/10.18273/revuin.v15n1-2016004>
- Calzada, L., & Abreu, J. L. (2009). El impacto de las herramientas de inteligencia de negocios en la toma de decisiones de los ejecutivos. *Revista Daena (International Journal of Good Conscience)*, 4(2). Recuperado a partir de

- [http://www.spentamexico.org/v4-n2/4\(2\)%2016-52.pdf](http://www.spentamexico.org/v4-n2/4(2)%2016-52.pdf)
- Camana, R. G. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica-ESPOL*, 29(1). Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/464>
- 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). Recuperado a partir de <http://www.redalyc.org/html/643/64316140004/CRISP-DM.pdf>. (s. f.).
- Flores, F. R., Pulido, L. F., & de la Rosa, E. D. (2016). Inteligencia de negocios y minería de datos aplicado a la industria refresquera. *Research in Computing Science*, 126, 63–71.
- Gallego, M., & Hernández, J. (2015). Identificación de factores que permitan potencializar el éxito de proyectos de desarrollo de software. *Scientia et Technica*, 20(1). Recuperado a partir de <http://www.redalyc.org/html/849/84938609009/>
- IBM Corp. (2012). Manual CRISP-DM de IBM SPSS Modeler. *IBM Redbooks*. Recuperado a partir de <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- IBM Knowledge Center. (2017). IBM SPSS Modeler CRISP-DM Guide. Recuperado 6 de octubre de 2017, a partir de [https://www.ibm.com/support/knowledgecenter/en/SS3RA7\\_17.1.0/modeler\\_crispdm\\_ddita/modeler\\_crispdm\\_ddita-gentopic1.html](https://www.ibm.com/support/knowledgecenter/en/SS3RA7_17.1.0/modeler_crispdm_ddita/modeler_crispdm_ddita-gentopic1.html)
- Jaramillo, A., & Arias, H. P. P. (2015). Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. *Revista Tecnológica-ESPOL*, 28(1). Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/351>
- Mariscal, G., Marbán, Ó., González, A., & Segovia, J. (2007). Hacia la Ingeniería de Data Mining: Un modelo de proceso para el desarrollo de proyectos. *Proceedings V Taller de Minería de Datos y Aprendizaje (TAMIDA'07)*. Pág, 139–148.
- Moine, J. M., Gordillo, S. E., & Haedo, A. S. (2011). Análisis comparativo de

- metodologías para la gestión de proyectos de minería de datos. En *XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011)*. Recuperado a partir de <http://sedici.unlp.edu.ar/handle/10915/18749>
- Montero, E., & Plasencia, A. (2016). Metodología CRISP-DM aplicada al proceso de minería de textos para agrupar documentos web. XIV Congreso Internacional de Información, Info'2016.
- Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the crisp-dm methodology. En *Proceedings of European Simulation and Modelling Conference-ESM'2011* (pp. 117–121). Eurosis. Recuperado a partir de <https://repositorium.sdum.uminho.pt/handle/1822/14838>
- Naranjo, R., & Sierra, L. (2009). Herramienta software para el análisis de canasta de mercado sin selección de candidatos. *REVISTA INGENIERÍA E INVESTIGACIÓN*, 29(1), 60-68.
- Rennolls, K., & Al-Shawabkeh, A. (2008). Formal structures for data mining, knowledge discovery and communication in a knowledge management environment. *Intelligent Data Analysis*, 12(2), 147–163.
- Rodríguez, M., Álvarez, V., Mesa, J., & González, A. (2003). Metodologías para la realización de proyectos de Data Mining. *AEIPRO Asociación Española de Dirección e Ingeniería de Proyectos*. Recuperado a partir de <https://sites.google.com/a/unicesar.edu.co/alvaroonate/pagina-1/parcelacion/METDOLOGIAS%20MINERIA%20DE%20DATOS.pdf>
- Rodríguez, R., & Cortés, F. (2012). Selección de una plataforma de inteligencia de negocios: un análisis multicriterio innovador. *Revista Ciencias Estratégicas*, 20(28). Recuperado a partir de <http://www.redalyc.org/html/1513/151326917003/>
- Romero, G., & Paredes, A. (2017). Análisis de la deserción estudiantil en la USB, facultad Ingeniería de Sistemas, con técnicas de minería de datos. *Revista Investigación y Desarrollo en TIC*, 4(1). Recuperado a partir de <http://publicaciones.unisimonbolivar.edu.co/rdigital/ojs/index.php/identific/article/view/1513>
- SAS Institute Inc. (2017, julio 3). Data Mining and SEMMA. Recuperado 29 de julio de 2017, a partir de <http://documentation.sas.com/?docsetId=emcs&docsetTarget=p0i38iu0h>

9sexjn1i3m2Int823xz.htm&docsetVersion=12.3&locale=en

- Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217–222.
- Timarán, S. R., Hernández, I., Caicedo, S. J., Hidalgo, A., & Alvarado, J. C. (2016). *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*. Universidad Cooperativa de Colombia. <https://doi.org/10.16925/9789587600490>
- Vanrell, J. A., Bertone, R., & García-Martínez, R. (2012). A Process Model for Data Mining Projects Un Modelo de Procesos para Proyectos de Explotación de Información. En *Proceedings Latin American Congress on Requirements Engineering & Software Testing LACREST* (p. 53). Recuperado a partir de <http://sistemas.unla.edu.ar/sistemas/gisi/papers/LACREST-2012-ISBN-978-958-46-0577-1-pag-46-52.pdf>
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. En *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29–39). Recuperado a partir de <https://pdfs.semanticscholar.org/48b9/293cfd4297f855867ca278f7069abc6a9c24.pdf>