



Universidad Nacional
Federico Villarreal

Vicerrectorado de
INVESTIGACIÓN

ESCUELA UNIVERSITARIA DE POST GRADO

“MODELO PARA LA APLICACIÓN DE MINERÍA DE DATOS ACADÉMICOS EN
LAS UNIVERSIDADES PERUANAS”

**TESIS PARA OPTAR EL GRADO ACADÉMICO DE:
DOCTOR EN INGENIERÍA DE SISTEMAS**

AUTOR:

HERRERA SALAZAR JOSÉ LUIS

ASESOR:

DR. GAMBOA CRUZADO JAVIER ARTURO

JURADO:

DR. SOTO SOTO LUIS

DR. HERNÁNDEZ CELIS DOMINGO

DR. RODRÍGUEZ RODRÍGUEZ CIRO

LIMA-PERÚ

2019

DEDICATORIA

Dedico mi trabajo de investigación a mi familia
que son la razón
de mi vida.

AGRADECIMIENTO

Agradezco a todas las personas que me apoyaron en la realización de la presente investigación en especial a mis amigos Jhony Pretell, Orlando Ypararraguirre, Luis Cuya, Ivan Petrlík y a mi Asesor de Tesis Dr. Javier Gamboa Cruzado.

RESUMEN

MODELO PARA LA APLICACIÓN DE MINERÍA DE DATOS ACADÉMICOS EN LAS UNIVERSIDADES PERUANAS

Herrera José Luis

La presente investigación ha partido de dos situaciones problemáticas: la mala gestión del proceso de minería de datos y la definición detallada de los entregables de cada etapa del proceso. Además de la inexistencia de un modelo de aplicación de Minería de datos en el entorno académico en las universidades peruanas.

Para desarrollar el Nuevo Modelo se revisaron metodologías robustas y modernas como: CRISP-DM, SEMMA, CATALYS, KDD y los aportes del investigador. Se hizo un análisis comparativo de las metodologías y se identificó que uno de los principales problemas en todas es la gestión en sí del proyecto por ende se utilizó el PMBOOK para mejorar esta etapa.

El nuevo modelo propuesto y aplicado consta de 2 subprocesos: Subproceso de Gestión de Proyecto, Subproceso de Aplicación de Minería de Datos. Así se logró la finalidad primordial que es Elaborar y aplicar el Modelo en una Institución Universitaria.

Excelentes resultados fueron obtenidos, entre ellos: El desarrollo del Nuevo Modelo para la mejora de proceso de aplicación de minería de datos en entornos académicos y en la aplicación del mismo donde se observó la optimización considerable de los indicadores propuestos.

Las conclusiones más importantes son: el nuevo modelo permitió la mejora en la gestión de proyectos de aplicación de minería de datos en entidades educativas universitarias, su aplicación permite obtener mejoras sustanciales al proceso, las técnicas estadísticas usadas durante las etapas fueron muy relevantes y significativas, y las TICs fueron herramientas relevantes que permitieron mejorar el proceso, para convertirlo en uno más eficiente y asegurar así que las tareas generen valor para los usuarios.

Palabras Clave: KDD, Minería de Datos, CRISP-DM, SEMMA, PMI, PMBOOK, Proceso.

ABSTRACT

MODEL FOR THE APPLICATION OF MINING OF ACADEMIC DATA IN THE PERUVIAN UNIVERSITIES

Herrera José Luis

The present investigation has started from two problematic situations: the bad management of the data mining process and the definition of the deliverables of each stage of the process. In addition to the lack of an application model of data mining in the academic environment in Peruvian universities.

To develop the New Model, robust and modern methodologies were reviewed such as: CRISP-DM, SEMMA, CATALYS, KDD and the contributions of the researcher. A comparative analysis of the methodologies was made and defined.

The new model proposed and applied consists of 2 subprocesses: Subprocess of Project Management, Subprocess of Application of Data Mining. This was achieved the primary purpose is to develop and apply the model in a University Institution.

Excellent results were seen, among them: The development of the new model for the improvement of the application process of data mining in the academic media and in the application of it where the proposed indicators can be considerably improved.

The most important conclusions are: the new model made possible the improvement in the management of the application projects of data mining in the university educational entities, its application to obtain substantial improvements in the process, the techniques of use of people. and ICT were tools to improve the process, to make it more efficient and ensure that tasks generate value for users.

Keywords: KDD, Data Mining, CRISP-DM, SEMMA, PMI, PMBOOK, Process.

ÍNDICE

DEDICATORIA.....	i
AGRADECIMIENTO.....	ii
RESUMEN.....	iii
ABSTRACT.....	iv
INTRODUCCIÓN.....	x
CAPITULO I: PLANTEAMIENTO DEL PROBLEMA.....	2
1.1. Descripción del Problema.....	2
1.2. Formulación de Problema.....	14
1.2.1. Problema general.....	14
1.2.2. Problemas específicos.....	14
1.3. Antecedentes.....	15
1.4. Justificación de la investigación.....	27
1.5. Limitaciones de la Investigación.....	28
1.6. Objetivos.....	28
1.6.1. Objetivo general.....	28
1.6.2. Objetivos específicos.....	28
1.7. Hipótesis.....	28
1.7.1. Hipótesis general.....	28
1.7.2. Hipótesis específicas.....	28
CAPITULO II: MARCO TEÓRICO.....	2
2.1. Marco Conceptual.....	30
2.1.1. Teorías generales.....	30
2.1.1.1. Minería de Datos.....	30
2.1.1.2. Analítica.....	31
2.1.2. Bases teóricas especializadas sobre el tema.....	35

2.1.2.1.	Minería de Datos Educativa (EDM) y Analítica de Aprendizaje (LA)	35
2.1.2.2.	Métodos de EDM.....	35
2.1.2.3.	Minería de datos para la educación.....	38
2.1.2.4.	Modelo de referencia para la Analítica del Aprendizaje	40
2.1.2.5.	Metodologías para análisis de datos	42
2.1.2.6.	Clasificación de técnicas de minería de datos utilizadas en entornos educativos	47
2.1.2.7.	Trabajos clasificados por técnicas de minerías de datos.....	47
2.1.2.8.	Clasificación de trabajos por dominios educativos.....	48
2.1.2.9.	Gestión y dirección de proyectos	50
2.1.2.10.	PMBOK	52
CAPITULO III: MÉTODO		30
3.1.	Tipo de Investigación.....	59
3.2.	Población y muestra.....	59
3.3.	Operacionalización de las variables.....	60
3.4.	Instrumentos.....	61
3.5.	Procedimiento	61
3.6.	Análisis de Datos	61
CAPITULO IV: DESARROLLO DEL NUEVO MODELO.....		62
4.1.	Elaboración del Modelo.....	63
4.1.1.	Representación Conceptual del nuevo modelo.....	64
4.1.2.	Roles	65
4.1.3.	Modelo propuesto de aplicación de Minería de Datos para el sector educativo.....	66
4.1.4.	Nivel de detalle de las actividades	67
4.1.5.	Actividades específicas de Minería de datos	69
4.1.6.	Actividades de dirección del proyecto.....	79
4.1.6.1.	Gestión del alcance	80

4.1.6.2. Gestión del tiempo	81
4.1.6.3. Gestión del costo.....	82
4.1.6.4. Gestión del equipo de trabajo	83
4.1.6.5. Dirección del riesgo	84
CAPITULO V: RESULTADOS	85
CAPÍTULO VI: DISCUSIÓN DE RESULTADOS	98
CAPÍTULO VII: CONCLUSIONES	102
CAPÍTULO VIII: RECOMENDACIONES.....	104
CAPÍTULO IX: REFERENCIAS	106
CAPÍTULO X: ANEXOS	113

ÍNDICE DE TABLAS

Tabla 1 <i>Preguntas cuyas respuestas son fuentes de información y conocimiento, mediante un análisis direccional</i>	33
Tabla 2 <i>Trabajos clasificados por técnicas de minería de datos</i>	48
Tabla 3 <i>Clasificación de trabajos por dominios presentes en la educación.</i>	50
Tabla 4 <i>Origen e historia del PMBOK</i>	53
Tabla 5 <i>Operacionalización de variables</i>	60
Tabla 6 <i>Roles de Gestión de Proyecto</i>	65
Tabla 7 <i>Características para el nivel de detalle de las actividades</i>	67
Tabla 8 <i>Evaluación de características con respecto a las fases del proceso de minería de datos</i>	69
Tabla 9 <i>Características de la Gestión del Alcance</i>	80
Tabla 10 <i>Características de la Gestión del Tiempo</i>	81
Tabla 11 <i>Características de la Gestión de Costos</i>	82
Tabla 12 <i>Característica de Gestión del equipo de trabajo</i>	83
Tabla 13 <i>Característica de la Dirección del Riesgo</i>	84
Tabla 14 <i>Resultados por cada indicador (KPI)</i>	86
Tabla 15 <i>Comparación de modelos a nivel de Modelo General</i>	99
Tabla 16 <i>Comparación de modelos a nivel del subproceso de Gestión de Proyectos</i>	100
Tabla 17 <i>Comparación de modelos a nivel del subproceso de Minería</i>	101

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Comparación de uso de metodologías extraído de (Gregory Piatetsky, 2014)	4
<i>Figura 2.</i> Izquierda: Modelo propuesto por Lee y Kerschberg (1998). Derecha: Modelo propuesto por Hofmann and Tierney (2009)	8
<i>Figura 3.</i> Ingresantes 2008 que se retiraron al 2012	13
<i>Figura 4.</i> Pasos que componen el proceso KDD (Fayyad, 1996)	30
<i>Figura 5.</i> Modelo de referencia para la analítica del aprendizaje, modificado de (Chatti et al., 2012).	41
<i>Figura 6.</i> Metodologías para análisis de datos adaptado de (IBM, 2012)	42
<i>Figura 7.</i> Semma.adaptado de (SAS Institute, 2018).....	43
<i>Figura 8.</i> P3QT	44
<i>Figura 9.</i> Framework para Minería de Datos Predictiva basado en el desarrollo de software adaptativo ASD	45
<i>Figura 10.</i> Modelo de Proceso KDDA en caracol	46
<i>Figura 11.</i> Diagrama de organización para modelos adaptado de Figuerola (2013) citado por Palacios (2014).....	51
<i>Figura 12.</i> Ciclo de vida del proyecto según costos y personal adaptado de Project Management Institute (2013).....	54
<i>Figura 13.</i> Grupo de Procesos de la Dirección de Proyectos adaptado de Project Management Institute (2013).....	55
<i>Figura 14.</i> Procesos del propuesto	64
<i>Figura 15.</i> Modelo propuesto.....	66
<i>Figura 16.</i> Estadística Descriptiva KPI1_PrePrueba	87
<i>Figura 17.</i> Estadística Descriptiva KPI1_PostPrueba.....	88
<i>Figura 18.</i> Estadística Descriptiva KPI2_PrePrueba	89
<i>Figura 19.</i> Estadística Descriptiva KPI2_PostPrueba.....	90
<i>Figura 20.</i> Estadística Descriptiva KPI3_PrePrueba	91
<i>Figura 21.</i> Estadística Descriptiva KPI3_PostPrueba.....	92

INTRODUCCIÓN

El objetivo fundamental en esta investigación es elaborar y aplicar un nuevo Modelo de Aplicación de Minería de Datos en entornos académicos apoyado con las TICs.

El punto de partida de la investigación tiene dos aspectos: el mal desempeño del proceso de minería de datos en instituciones académicas universitarias; y el problema de la gestión de proyectos en los modelos existente de aplicación de minería de datos.

El respaldo teórico y filosófico de la Investigación es: Modelos de Aplicación de Minería de Datos, Minería de Datos en el entorno educativo, Gestión de proyectos bajo el enfoque PMBOOK . Se demuestra que al aplicar el nuevo modelo este cubre las falencias del resto de modelos en la fase de gestión del proyecto; también se observa la mejora de los valores de los KPIs del proceso de aplicación de minería de datos en una Institución Universitaria.

El nuevo modelo se realizó aplicando una combinación de los modelos CRISP-DM y el marco de trabajo del PMBOOK generando 2 subprocesos: Subproceso Gestión de Proyecto y Subproceso Aplicación de minería de datos.

También se utilizan TICs relevantes como Bizagi y MiniTab.

Se encontraron algunas limitaciones en la investigación: falta de tiempo del investigador y algunas restricciones en la información institucional.

Para que sea más clara la presente tesis, se ha dividido en seis capítulos:

El **Capítulo I: Planteamiento del Problema**, detalla los antecedentes ubicados en tesis, libros y artículos científicos, la Realidad Problemática, Objetivo General y Específicos, Justificación, Limitaciones.

En el **Capítulo II: Marco Teórico**, se detallan, las bases teóricas referentes a las metodologías utilizadas en el desarrollo de la investigación. Además del fundamento filosófico de la investigación.

El **Capítulo III: Método**, describe el Tipo y Nivel de la Investigación, las Hipótesis, sus Variables, Diseño de Experimentos y las Técnicas e Instrumentos para el Registro de Datos.

El **Capítulo IV: Desarrollo del nuevo modelo** es la parte medular de la investigación ya que muestra detalladamente la creación y aplicación de un nuevo modelo para la mejora de procesos adecuada al sector educativo universitario. Se aplica el nuevo modelo en el desarrollo de un proyecto de Minería de Datos para determinar los estilos de aprendizaje de los estudiantes universitarios.

En Resultados y Contrastación de la Hipótesis desarrollado en el **Capítulo V y VI**, se inicia identificando la Población y la Muestra, Análisis e Interpretación de Resultados. Finalmente se desarrolla la validación de las Hipótesis.

Luego se tiene el **Capítulo VII, VIII y IX**, donde se realiza la Discusión de resultados, las Conclusiones y Recomendaciones.

Finalmente se detallan las Referencia Bibliográficas y Anexos

CAPITULO I:
PLANTEAMIENTO DEL
PROBLEMA

1.1. Descripción del Problema

En el proceso de Aplicación:

Los orígenes del *Data Mining* se desarrollaron a inicios de los 80s, cuando la gestión de hacienda estadounidense desarrolló un programa de investigación para descubrir fraudes en la afirmación y evasión de impuestos, por medio de lógica difusa, redes neuronales y técnicas de reconocimiento. El enorme avance tecnológico de las PCs en el último período ha maximizado el alojamiento de enormes proporciones de datos y permitió el avance de utilidades para su régimen, dando lugar a una exclusiva especialidad popular como “data mining”.

No obstante, la enorme propagación del *Data Mining* no se produce hasta los 90s originada primordialmente por tres factores:

- Incremento de la capacidad de los ordenadores
- Incremento del ritmo de compra de datos. El desarrollo de la proporción de datos guardados se ve favorecido no sólo por el abaratamiento de los discos y sistemas de alojamiento masivo, sino además por la automatización de varios trabajos y técnicas de obtenida de datos.
- Aparición de nuevos procedimientos de técnicas de estudio y alojamiento de datos.

Lamentablemente esta propagación supone el avance de proyectos cada vez más importantes en un área en el que difícilmente se tienen la posibilidad de obtener conclusiones a priori y en el que la selección de la preferible técnica es imposible llevar a cabo en las primeras fases, sino que se precisa un modelo evolutivo, semejante al modelo espiral del período de vida de avance programa. Por otro lado, visto que más del 75% del esfuerzo se genere en las primeras fases (en esta situación en el pretratamiento de datos) hace que esta clase de proyectos sea generalmente subestimado en relación a coste y tiempo y que las desviaciones producidas superen con bastante el 90%.

No obstante, y más allá de que hoy en día se cuenta utilidades de procesamiento capaces y a lo mejor las restricciones, en relación a infraestructura no se muestran como en los inicios del Dataminig. Se ha constatado que varios de estos proyectos no acaban y que, inclusive habiendo terminado, éstos no lo hacen en los plazos y/o con los capitales previstos o no corresponden con las expectativas de los usuarios. Entre las primordiales causas

identificadas y que comentan estos sucesos, están las similares con la carencia de procesos de avance estandarizados que incorporen un enfoque ingenieril al avance de proyectos de Data Mining (Gallardo Arancibia, 2009)

Frente la necesidad que existe en el mercado de una proximidad sistemática para la ejecución de los proyectos de Data *Mining*, distintas compañías y consultorías han especificado un desarrollo de modelado pensado para asesorar al usuario por medio de una serie de pasos que le dirijan a conseguir excelente resultados. De esta forma SAS ofrece la utilización de la metodología SEMMA (Sample, Explore, Modify, Model, Assess). En 1999 un considerable consorcio de compañías de europa, NCR (Dinamarca), AG(Alemania), SPSS (Inglaterra) y OHRA (Holanda), unieron sus elementos para el avance de la metodología de libre organización CRISP-DM (Cross-Industry Standard Process for Data Mining). Esta metodología, adjuntado con la metodología SEMMA, son ámbas primordiales formas usadas por los investigadores en los proyectos de Data Mining

Las investigaciones llevadas a cabo por kdnuggets entre 2002 y 2014 tienden a reflejar el dominio de CRISP-DM como metodología más empleada en la administración de proyectos de Data Mining, seguida de SEMMA y KDD, sin embargo, hay un porcentaje sustancial de profesionales que utiliza su propia metodología, una establecida por la organización donde trabaja u otras formas de menor calado.

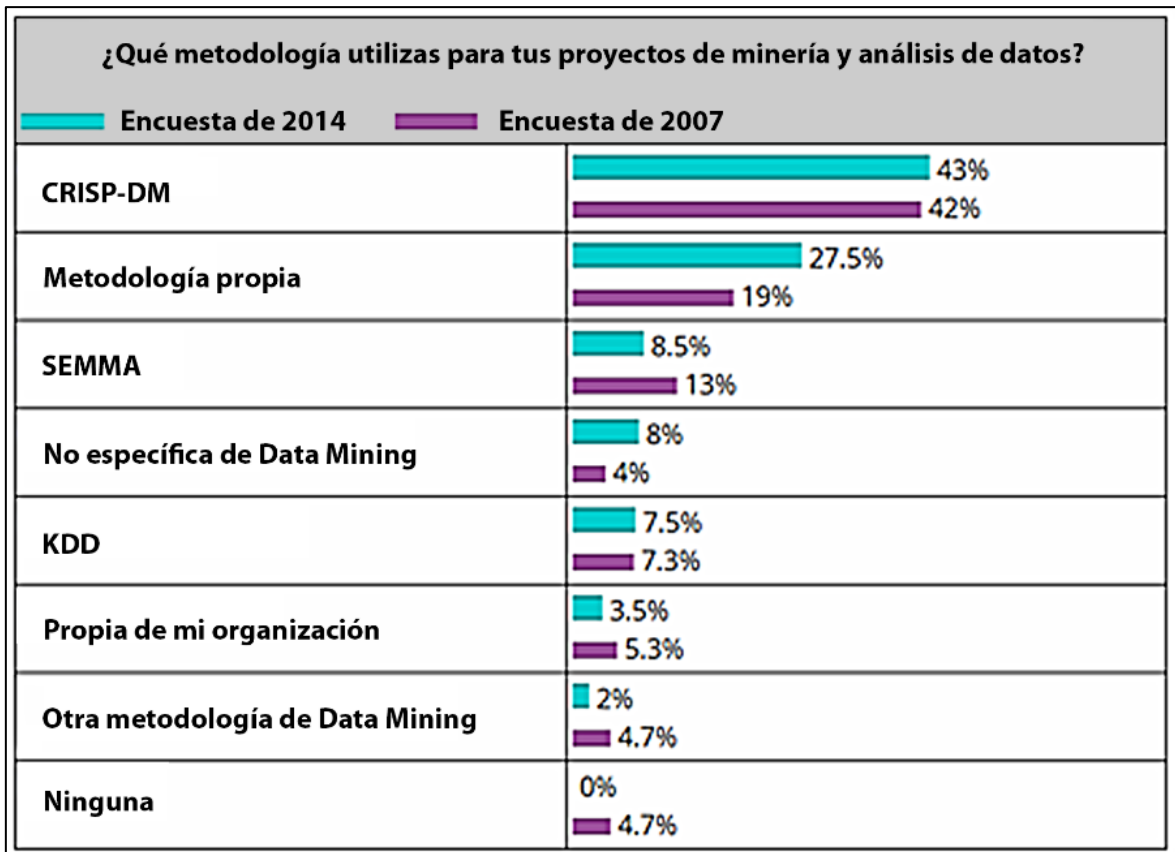


Figura 1. Comparación de uso de metodologías extraído de (Gregory Piatetsky, 2014)

Según (Marbán, Mariscal, Menasalvas, & Segovia, 2007), se sugiere que la tasa de fracaso en proyectos correspondientes a la especialidad sobrepasa el 60%, siendo CRISPDM la metodología más usada. Desde el avance de diversos proyectos de explotación de información, aplicando los diferentes modelos de procesos de explotación de información que ya están, se identifican una secuencia de deficiencias y carencias para la ejecución triunfadora del emprendimiento.

Entre las primordiales causas ciertas que argumentan la pregunta, ¿por qué varios de los proyectos de Data Mining, no concluyen con triunfo?, están las similares con la carencia de procesos de avance estandarizados que incorporen un enfoque ingenieril al avance de proyectos de Data Mining y puntos relacionados con una indebida etapa de especificación de requisitos.

La falla puede suceder por muchas causas, no obstante, hay algunos peligros evidentes que van a hacer que algún emprendimiento de big data se bloquee y se queme. Basado en mi vivencia haciendo un trabajo con compañías y organizaciones de todas las formas y tamaños, sé que estos fallos son muy recurrentes. Una cosa que tienen parecido es que todos son causados por la carencia de una idealización correcta. (Bernard, 2015)

Los proyectos de Explotación de Información son un tipo particular de emprendimiento de Ingeniería en Software. En vez de necesitar desarrollar un programa concreto, utilidades accesibles son usadas que ya tienen dentro las técnicas y algoritmos necesarios. Como resultado las propiedades de los proyectos de Explotación de Información son distintos a los de la Ingeniería en Programa Clásico y de la Ingeniería del Conocimiento. Pero de cualquier manera tiene inconvenientes semejantes. Estudios completados sobre sobre proyectos de Explotación de Información detectaron que la mayor parte de los proyectos finaliza en fracaso. En el año 2000 se ha había preciso que el 85% de los proyectos no alcanzan sus misiones, en tanto que en el 2005 el porcentaje de fracaso bajo a precisamente el 60%. Entonces, se puede decir que la red social estuvo haciendo un trabajo en el sendero acertado, pero hay cuestiones de administración que aún tienen que ser mejorados. (García-Martínez, R., et al., 2011)

De la misma manera que los modelos de desarrollo de programa clásico, en sus primeras etapas se enfocaban en las ocupaciones asociadas a la algoritmia y programación del programa (Marbán et al., 2007), los modelos de procesos de explotación de información, se centran en su actividad semejante (Minería de datos).

El modelo de desarrollo KDD tiene como primordial carencia la carencia de información es en relación a las ocupaciones primordiales para gestionar un emprendimiento de explotación de información. CRISP-DM más allá de que identifica algunas escasas ocupaciones enlazadas a la gestión de un emprendimiento, este está focalizado en el avance de las ocupaciones técnicas del emprendimiento, sin tener en cuenta puntos como medidas, estimación, tercerización, administración del cambio, etc. (Marbán et al., 2007)

Complementariamente, los modelos de procesos nombrados, muestran deficiencias en la estructuración de las tareas en general provocando iteraciones innecesarias que causan demoras y aumentos en los costos. La elaboración de los datos antes de determinar las utilidades y algoritmos a usar implica a una viable necesidad de iterar de nuevo hacia la etapa de elaboración de los datos, siendo viable detectar estos elementos de forma previa. La selección de los algoritmos de minería de datos a usar es dejada a la experticia del ingeniero de minería de datos. Se omiten las vinculaciones y dependencias entre los elementos producidos en todo el desarrollo.

Entre los autores que piensan necesaria la construcción de una exclusiva metodología para la administración de proyectos de minería de datos está (Saltz, 2015). Según el señor Saltz el aumento de la agilidad, volumen y diversidad de datos accesibles para las organizaciones ha incrementado el interés y el número de proyectos relacionados con el aprovechamiento de estos datos. Esto ha derivado en equipos cada vez más importantes, los cuáles involucran cada vez aún más grande número de expertos, en muchas oportunidades de diferentes disciplinas. Saltz piensa que las formas habituales no se han refinado con el propósito de hacer más simple equipos multidisciplinarios o de enorme amplitud si no que los expertos del área se han centrado solamente en la optimización de las técnicas de extracción y examen de los datos. Un apunte que exhibe esta falta de interés es que de los 296 artículos e itinerarios de charlas distribuidos a lo largo de la “Conferencia sobre Big Data 2014” promovida por IEEE (Institute of Electrical and Electronics Engineers) ninguno se centraba en actualizaciones sobre las formas de administración de proyectos en el sector y solo un 8% mencionaba los retos técnicos y sociales que puede llegar a implicar un emprendimiento de minería de datos.

Otro aspecto relevante mencionado por Saltz es que generalmente las metodologías de minería de datos suelen tomar una aproximación de “tareas” hacia el problema, conformado por una serie de etapas que pueden repetirse de forma iterativa. Este es un punto de vista que no ha evolucionado, en su opinión, en al menos 20 años, siendo la mayoría de las metodologías adaptaciones de KDD y solo diferenciándose algo CRISP-DM, la cual valora como un primer paso hacia el establecimiento de una metodología completa para la gestión de estos proyectos.

Sus motivos para considerar necesaria la definición de una nueva metodología son los siguientes:

1. La descripción paso por paso empleada en las metodologías tradicionales no está pensada para grandes equipos. Por ello se acaban realizando soluciones específicas para cada tipo de proyecto no siendo generalizables y debiendo llegar a ellas por el método prueba-error. Esto denota una “baja madurez” del sector.
2. Existe cierta dificultad de prever si el proyecto va a ser exitoso, se va a terminar su ejecución en tiempo y dentro de presupuesto.

3. Hace falta una metodología que defina con claridad la organización del grupo de trabajo, las funciones de cada individuo o el control de costes y plazos.

De acuerdo con Saltz la metodología propuesta debe de abordar al menos los siguientes problemas:

- Coordinación del equipo
- Gestión de la calidad de los resultados
- Propiedad intelectual de los datos, seguridad y privacidad
- Análisis de requisitos
- Priorización de requisitos
- Implantación

Por otro lado, Javier Segovia establece también un punto muy válido. En su opinión CRISP-DM intenta ser al mismo tiempo un modelo de proceso, una metodología y un ciclo de trabajo, esto hace que acabe careciendo de definición y detalle. En la opinión de Segovia esto hace que el proceso de minería de datos se encuentre aislado del resto del proceso de ingeniería de software pese a que tiene una implicación muy importante en el mismo. Si bien Segovia piensa en un proyecto puramente de software al hacer esta reflexión esta misma es bien aplicable a otros tipos de proyectos. Las metodologías tradicionales de minería de datos entienden, tal y como mencionó Saltz los proyectos como una serie de pasos sucesivos los cuáles llevan o bien a la consecución de un objetivo o deben de ser repetidos hasta hallar el resultado deseado. Esto produce de algún modo una abstracción del resto de variables y entorno que rodean el proyecto (competidores, involucrados, mercado, ambiente dentro de la organización ...) (Pérez, 2007)

En los últimos años ha habido un buen número de autores que han propuesto metodologías alternativas a SCRUM-DM y el resto de metodologías de gestión de proyectos de minería de datos con el fin de mejorar la eficiencia en la aplicación de estas técnicas a algún campo o área en particular. No obstante, la mayoría de estas metodologías difieren poco de las metodologías tradicionales, limitándose a una mera reinterpretación de alguna de las fases que componen las mismas con el fin de facilitar su adaptación a un campo u organización.

Ejemplos de estas son:

- La reinterpretación de Robin Way de la metodología SEMMA, ligada altamente al sector bancario y de inversiones y con un fuerte énfasis en la fase de implantación. (Robin Way, 2013)
- La adaptación de CRISP-DM al sector médico de Olegas Niakšu, quien establece una descomposición de las fases que componen la metodología CRISP en procesos específicos para el sector médico. (NIAKŠU, 2015)

Lee and Kerschberg quienes establecen un procedimiento de control para la metodología KDD mediante un panel de expertos. (Lee & Kerschberg, 1998)

Hofmann and Tierney quienes establecen responsables para cada una de las etapas del KDD de 9 pasos. (Hofmann & Tierney, 2009)

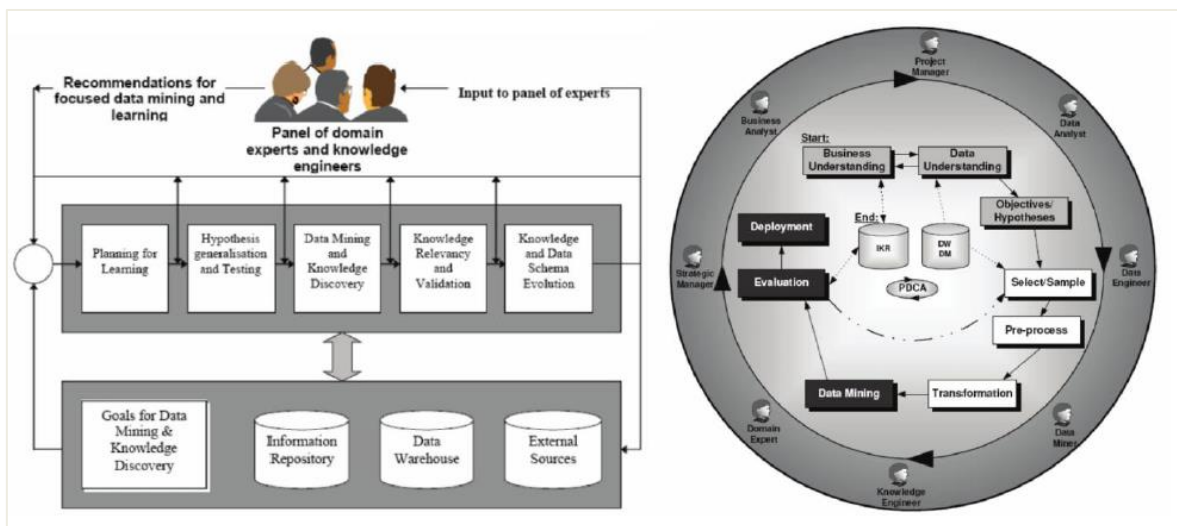


Figura 2. Izquierda: Modelo propuesto por Lee y Kerschberg (1998). Derecha: Modelo propuesto por Hofmann and Tierney (2009)

Estas formas, aunque atractivas no aportan ideas recientes y se limitan a ser adaptaciones de las formas que ya están para un fin en especial. Sin embargo, a lo largo del estudio del “Estado del Arte” se ha detectado unos pocos autores que se considera aportan ideas recientes en la materia y cuyas proposiciones se cree se distinguen lo bastante de las formas habituales de administración de proyectos de minería de datos. De las insuficiencias identificadas aparece como inconveniente la necesidad de desarrollar un modelo de desarrollo de avance de proyectos de ingeniería de explotación de información que guíe de manera precisa tanto en las cuestiones técnicas del emprendimiento, de esta forma como las de administración del mismo, siendo primordial para eso detectar las fases, ocupaciones,

tareas, técnicas de representación y métodos de ejecución los cuales permitan sistematizar el avance de proyectos en el sector.

En la Educación Universitaria:

“La noción de sociedad de la información se basa en los progresos tecnológicos. En cambio, el concepto de sociedades del conocimiento comprende dimensiones sociales, éticas y políticas mucho más vastas. El hecho de referirse a sociedades, en plural, no se debe al azar, sino a la intención de rechazar la unicidad de un modelo “listo para su uso” que no tenga suficientemente en cuenta la diversidad cultural y lingüística, único elemento que permite a cualquiera reconocerse en los cambios que se están produciendo actualmente” (I. M. d. I. UNESCO, 2005).

Hoy en día, uno de las formas más comunes de archivar/almacenar la información no solo educativa sino personal es en formato digital. Esto crea un vasto acervo de conocimientos que solo se puede acceder, por razones obvias, a través de herramientas digitales. Los usuarios de esta información tendrán la fluidez en las herramientas de acceso digital, exploración, visualización, análisis y colaboración. Estas herramientas influyen tanto en la creación y el análisis de la información.

Los analistas basan su proceso de producción de los datos, hipótesis y las preguntas a resolver dependiendo de las posibilidades que sus herramientas les ofrecen. El almacenamiento digital y la interconectividad de la Web ofrecen nuevos retos en términos de cantidad y calidad de la información. Crean un nuevo medio para la presentación, así como una base para la colaboración, que es independiente de la ubicación física.

La información en forma digital ofrece oportunidad sin igual para combinar, extraer, presente y compartir ideas complejas. Existen grandes retos en la era del conocimiento: el desarrollo de nuevos esquemas y paradigmas para la representación de la compleja y vasta información actual que puede ser compartida, analizada y comparada; la creación de una alfabetización en el análisis de información y visualización que tiene el mismo rigor y la riqueza como la actual; y utilizar eficazmente la inmensa cantidad de datos obtenidos, computados y almacenados por los sistemas de información actuales, con el propósito de obtener nuevo conocimiento.

Este panorama es común a prácticamente todas las actividades que desarrollan las personas en sociedad, y los procesos de aprendizaje están cada vez más inmersos en esta vorágine

generadora de datos. En este contexto, el razonamiento analítico facilitado por las interfaces visuales interactivas y el uso de métodos computacionales adecuados se revelan como herramientas indispensables.

Actualmente, con la introducción de las TIC en la educación, se ha fomentado el uso de la tecnología con fines de apoyo al proceso de aprendizaje y enseñanza. Este fenómeno, junto con la aplicación de nuevas técnicas de procesamiento de los datos, que se han ido adaptando desde otros campos (por ejemplo el empresarial), han propiciado la necesidad de un análisis de datos avanzado en la educación, dando origen al término analítica académica (Philip J. Goldstein & Richard N. Katz, 2005).

La Analítica del Aprendizaje o Analítica Educativa (del inglés, Learning Analytics, LA) responde a la necesidad de llevar a cabo el seguimiento y control de la actividad en el campus para la toma de decisiones estratégicas, mediante el aprovechamiento de la gran cantidad de datos producidos por los estudiantes en actividades académicas (Fournier et al., 2011).

La sociedad se encuentra en la denominada era de la información, los datos almacenados se incrementan significativamente en diferentes fuentes, estructuras y formatos; pero, contrariamente esta expansión no siempre supone un aumento de conocimiento, debido a que procesarlos con los métodos clásicos resulta sumamente tedioso y con resultados superficiales e insatisfactorios. Nos enfrentamos a la paradoja de que, cuantos más datos están disponibles, menos conocimiento se tiene.

La solución evidentemente está en tener la información adecuada, en el lugar y momento oportuno, y así incrementar la efectividad de las universidades. La idea clave es que los datos contienen más información oculta de la que se ve a simple vista, por lo que hay que “torturarlos hasta que ellos confiesen” (Molina Félix, 2002), que es una explicación informal de la actividad que se realiza mediante la, denominada Minería de Datos (MD). La Minería de Datos (MD), por ejemplo, técnicas, uso de inteligencia Artificial para hallar patrones y relaciones entre los datos, admitiendo la construcción de modelos y representaciones abstractas de la verdad. De esta forma el valor real de los datos radica en la información que se puede obtener de ellos, información que ayude a tomar elecciones o hacer mejor nuestra comprensión de los fenómenos que nos cubren. Crear conocimiento es el desarrollo no trivial de detectar patrones válidos, noticiosos, probablemente servibles y, en más reciente instancia, comprensibles desde los datos importantes y nuevos sobre un fenómeno o

actividad por medio de algoritmos eficaces, dadas las crecientes órdenes de intensidad en los datos. (Valcárcel Asencios, 2004).

Con toda esa información que poseen las universidades las metas son: procesar automáticamente esas grandes cantidades de datos crudos, identificar los patrones más significativos y relevantes, presentarlos como conocimiento apropiado para satisfacer las demandas de la gestión educativa. Este conocimiento obtenido permitirá llevar a cabo acciones, ya sea incorporándolo dentro de un sistema de informático o para almacenarlo y reportarlo a los interesados, en este sentido, implica un proceso interactivo e incremental.

Las universidades coleccionan datos y se espera que de ellos emerjan hipótesis como por ejemplo porque los estudiantes desertan, porque los estudiantes salen desaprobados, cuáles son los estilos de aprendizaje predominante, en ese sentido se busca que los datos describan o indiquen por qué son como son para luego entonces, validar esas hipótesis; será numéricamente significativa, pero experimentalmente inválida. De ahí que la minería de datos debe presentar un enfoque exploratorio, y no confirmador usarla para confirmar las hipótesis formuladas puede ser peligroso, pues se está haciendo una inferencia poco válida. La deserción, el bajo rendimiento y los bajos índices de eficiencia terminal se encuentran entre los problemas frecuentes más complejos de la gestión académica que enfrentan las universidades. En ese sentido ha emergido la Minería de Datos Educativa (EDM) quien se preocupa por desarrollar métodos para explorar las características singulares de los datos que provienen de entornos educativos y utilizar esos métodos para comprender mejor el desempeño de los estudiantes, y las condiciones en las cuales ellos aprenden. Es el proceso de transformar los datos en bruto, recopilados por los sistemas de enseñanza, en información útil que pueda utilizarse para tomar mejores decisiones y que respondan preguntas de investigación.

Según el informe de RENAME: Registro Nacional de Alumnos, Maestros y Escuelas de Mexico (Bataller Sala, 2010) los sistemas educativos tienen los siguientes problemas con respecto a la información:

- Falta de capacidad institucional para elaborar diagnósticos que permitan hacer una buena planeación, anticiparse a las necesidades del sistema educativo y facilitar el consenso nacional de políticas educativas.
- No existen condiciones para generar información suficiente, actualizada, oportuna, confiable y más útil para la toma de decisiones.

La educación superior en América Latina presenta altas tasas de deserción estudiantil, especialmente en los primeros semestres académicos, lo cual conlleva a efectos de tipo financiero, académico y social, tanto para las Instituciones de Educación Superior (IES) como para el estudiante, la región, el país y el Estado. Según el Instituto para la Educación Superior en América Latina y el Caribe (IESALC), Latinoamérica presentó en el año 2003 una cobertura promedio en educación superior del 28.7%, y una tasa de deserción estudiantil del 50%. (Timarán Pereira, Calderón Romero, & Jiménez Toledo, 2013).

Las universidades peruanas pasan por varios problemas uno de ellos es la Deserción, las causas que llevan a un estudiante a desertar son diversas, son pocos los estudios realizados sobre Deserción ya que es difícil encontrar al desertor, pues no se encuentra en la universidad, pero para cubrir ésta falta de información las universidades realizan un seguimiento a sus estudiantes, considerándose desertor a aquel estudiante que en tres años consecutivos no haya registrado matrícula. Vemos también que, entre las universidades informantes, las universidades públicas son las que tienen un mayor porcentaje de deserción (27,8%), porcentaje que no es muy lejano al de las universidades privadas (20,5%). (ANR, 2014)

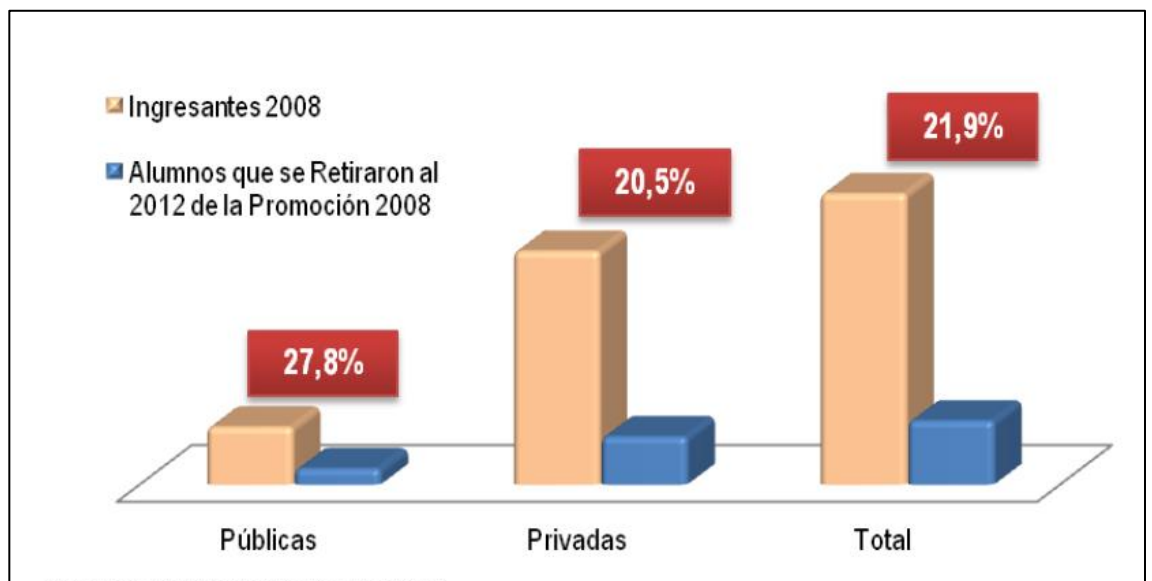


Figura 3. Ingresantes 2008 que se retiraron al 2012

Fuente: Universidades del Perú

La aplicación de la Minería de Datos en el ámbito de la enseñanza, tiene como objetivo obtener una mejor comprensión del proceso de aprendizaje de los estudiantes y de su participación global en el proceso, orientado a la mejora de la calidad y la eficiencia del sistema educativo (Jiménez Galindo & Álvarez García, Minería de Datos en la Educación, 2010).

Es así que, a partir de toda la información disponible, las diferentes técnicas de Minería de Datos pueden ser aplicadas a fin de descubrir conocimiento útil que ayude a mejorar la gestión educativa de las universidades, siendo este conocimiento muy diverso. Puede estar dirigida a los alumnos, profesores o autoridades académicas, quienes a partir de ella pueden identificar tres tipos de objetivos: pedagógicos (ayuda en el diseño de contenidos didácticos, mejoras en el rendimiento académico de los alumnos), de gestión (optimizar la organización y mantenimiento de infraestructuras educativas, áreas de interés, cursos más solicitados) y comerciales (permite realizar segmentación del mercado y facilita la captura de alumnos) (Gómez, García, & Therón, 2014)

1.2. Formulación de Problema

1.2.1. Problema general

¿De qué manera el desarrollo de un modelo MoPro-DM permitirá mejorar el proceso de aplicación de minería de datos académicos en las universidades peruanas en comparación con otros modelos?

1.2.2. Problemas específicos

- ¿En qué medida el modelo mejorará Porcentaje de Cumplimiento Metodológico?
- ¿En qué medida el modelo mejorará el Porcentaje de Cumplimiento de la Gestión de Proyecto?
- ¿De qué manera el modelo mejorará Porcentaje de cumplimiento en actividades de Desarrollo de MD?

1.3.Antecedentes

1) Autor: Jesús Germán Andrés Pautsch

Título: Minería de Datos Aplicada al Análisis de la Deserción en la Carrera de Analista en Sistemas de Computación.

Resumen:

En los años anteriores, la tasa mundial aumentó en el área informático al 20% anual, en tanto que en Argentina fue menor al 8 % anual. Este porcentaje tiene la posibilidad de ser superior si egresaran de las universidades más expertos informáticos, más que nada de las carreras a corto período que son las que las compañías superiores denuncian. En esta exploración, se procedió a hacer la Minería de Datos sobre el Cubo 04 Desgranamiento, exportado del Sistema de Administración Académica SIU-Guaraní, proporcionados por el Ministerio de Educación, Ciencia y Tecnología de la Nación. Se ha experimentado con procedimientos Supervisados, como la Categorización y No Supervisados como es el Agrupamiento. El propósito primordial fue incrementar la calidad de los modelos para clasificar y agrupar a los alumnos, según sus propiedades académicas, componentes sociales y demográficos, que han desertado de la Carrera Analista en Sistemas de Computación de la Facultad de Ciencias Exactas, Químicas y Naturales de la Facultad Nacional de Misiones. La proposición se desarrolló bajo la metodología de libre difusión Crisp-DM, y con la utilidad comercial IBM DB2 Warehouse (versión 9.5). Los resultados que se consiguieron permitieron ver, no sólo los distintos patrones de los estudiantes desertores, sino que además las propiedades de los alumnos en actividad y egresados. La calidad de los modelos obtenidos por medio de la categorización con árboles de elección superó a la técnica de agrupamiento por medio de la generación de clústeres y las dos han superado extensamente lo planteado. Más allá de que los profesionales en el estudio de la deserción universitaria han elogiado la aptitud que los modelos tienen para distinguir a cada tipo de alumno, han planteado la necesidad de contemplar más cambiantes colegas económicas en el estudio de la deserción. (Andrés Pautsch , 2009)

2) Autores: Alex Fernando Bonilla Gordillo, Miguel Ángel Ojeda Schuldt

Fabricio Echevarría Briones

Título: Implementación de minería de datos basada en redes bayesianas para la toma de decisiones en los registros académicos.

Resumen:

En el presente informe no explica cómo sería el sistema de un colegio implementando la minería de datos basadas en redes bayesianas la cual provee una forma compacta de representar el conocimiento y métodos flexibles de razonamiento basados en teorías probabilísticas la cual es llamada Teorema de Bayes que son capaces de predecir el valor de las variables no observadas y explicar las observadas.

Antes de abordar con el tema se explicó en el trabajo diversas metodologías que no cumplían algunos requerimientos como el caso de la metodología OLAP la cual no generaba patrones, reglas ni pautas lo cual era necesario para diagnosticar fácilmente los registros académicos y de ello poder tomar decisiones.

Las soluciones que presentan son probabilidades de eventos que se relacionan a las situaciones deseadas; la cual en la pregunta debe existir la interrogante y la variable a cuestionar, por ejemplo ¿Cuál sería la probabilidad de que alguien se registre en la materia X? al agregarse un mayor tipo de condiciones las respuestas se harán más específica y podrá filtrar las variables de registro.

Aplicando las redes bayesianas se puede dar soporte para la toma de decisión y así reducir uno de los problemas más comunes al momento de un registro en paralelo, como podría serlo en registro vacíos donde no hay datos de donde buscar. (Bonilla Gordillo, Ojeda Schuldt, & Echeverria Briones, 2009)

3) Autor: Fing Luan, PhD

Título: Aplicaciones de Minería de Datos en la Educación Superior

Resumen:

Uno de los superiores desafíos a los que se enfrenta la educación superior, en la actualidad es el pronóstico de las trayectorias particulares de los alumnos y de los antiguos estudiantes. A las instituciones académicas les agradecería entender, entre otras cosas, qué estudiante va a necesitar asistencia para graduarse, qué estudiante se matriculará en un programa concreto, o qué ex-alumno, está más abierto a qué tipo de operación de donaciones, Además, los inconvenientes recurrentes como la administración de las matrículas y el tiempo que demora un estudiante en terminar sus estudios, siguen motivando a las instituciones de educación superior para buscar superiores resoluciones. Una manera de enfrentar estos desafíos, es por medio de la minería de enormes almacenes de datos presentes en los campus universitarios. La minería de datos facilita a las organizaciones conocer y comprender los patrones ocultos en enormes bases de datos. Ahora, estos patrones se integran en modelos de minería de datos y se usan para pronosticar las formas de proceder particulares con enorme exactitud. Entre otras cosas, la minería de datos puede prestar a una institución académica la información que se requiere para tomar medidas antes de que un estudiante abandone, o para conceder de forma eficiente los elementos a una estimación precisa del número de alumnos que se matricularán en un curso preciso. Como producto de esta clase de información, las instituciones académicas tienen la posibilidad de conceder al personal y los elementos de forma más eficaz. Este archivo técnico trata de las funcionalidades de la minería de datos y de las apps en la educación superior. Tres estudios de casos comprueban cómo la minería de datos ahorra los elementos, mientras capacidad totalmente la eficacia e aumenta la eficacia sin que aumente el coste. El archivo comienza con una especificación general de las funcionalidades de la minería de datos.(Jing , 2010)

4) Autor: Álvaro Jiménez Galindo Hugo Álvarez García

Título: Minería de Datos en la Educación

Resumen:

La minería de datos educacionales da múltiples virtudes, comparándola con los paradigmas habituales de exploración relativa a la educación, como experimentos de laboratorio, estudios sociológicos o exploración de interfaz. La construcción de repositorios públicos de datos educacionales, creó una base que hace viable la minería de datos. En especial, los datos de estos repositorios son completamente válidos (son datos reales sobre el desempeño y estudio de alumnos, en ambientes educacionales, tomados en tareas de aprendizaje), y cada vez más de forma sencilla accesibles para empezar una exploración. Dando permiso a los estudiosos ahorrar más grande tiempo en tareas como la búsqueda de individuos (como academias, instructores y alumnos), organización de los estudios y recolección de datos, dado que estos están de manera directa accesibles.

Aunque la utilización de datos antes recogidos limita los exámenes a las cuestiones que conciernen a estos, una exploración previa puede ser increíblemente servible para investigar cuestiones poco similares con los datos tomados, como, entre otras cosas atributos de los alumnos así como razón o accionar estratégico. La disponibilidad de estos datos ha supuesto un enorme avance. Una vez definido un modelo de interés educativo, puede probarse con nuevos conjuntos de datos. La transferencia de estos modelos puede no ser trivial, pero el desarrollo de avance y validación de un modelo para un nuevo contexto es bastante más acelerado. Además, la presencia de una cantidad enorme de estudiantes que utilizan utilidades de estudio semejantes, aunque sea en diferentes contextos, contribuye una oportunidad novedosa de estudiar la predominación de componentes contextuales en instructores y estudiantes. Históricamente, fué muy complicado estudiar cómo las diferencias entre grupos de instructores, o clases influyen en puntos particulares del estudio. Esta clase de exámen es mucho más simple con la minería de datos. De forma semejante, el encontronazo de diferencias particulares fué difícil de estudiar estadísticamente con procedimientos habituales. La minería de datos aplicada al ámbito educativo tiene el potencial de prolongar un grupo de utilidades más extenso para el exámen de cuestiones sobre diferencias particulares. (Jiménez Galindo & Álvarez García, Minería de Datos en la Educación, 2010)

5) Autor: Karina Eckert, Roberto Suénaga

Título: Aplicación de técnicas de Minería de Datos al análisis de situación y comportamiento académico de alumnos de la UGD

Resumen:

En el tema educativo es visible la necesidad de tener sistemas de administración que permitan tomar elecciones académicas y llevar a cabo tácticas desde el conocimiento oportuno, dado que esto no solo influye de manera directa sobre la utilidad de los departamentos académicos, u otras cuestiones internas; sino que además podrían incidir sobre ocupaciones como las evaluaciones y acreditaciones de instituciones y carreras. Entre los inconvenientes menos simples que combaten las instituciones educativas, tenemos la posibilidad de mencionar: hacer mejor la calidad académica, achicar la deserción y la reprobación, evadir el atraso académico y los bajos índices de eficacia relacionado con las tasas de graduación. Esto necesita administrar tácticas y tomar medidas frente a estos acontecimientos; para eso es viable recurrir al desarrollo llamado Minería de Datos Educacional (MDE). La aplicación del desarrollo de Hallazgo o Extracción de Conocimiento en Bases de Datos (KDD) en tema educativo. En la presente exploración, se detalla y muestra la aplicación del desarrollo KDD (por su sigla en inglés), popular como Minería de Datos (MD) en un ámbito educativo, más exactamente a la información académica de la Facultad Gastón Dachary (UGD). El desarrollo radica en una sucesión de etapas que parten de la selección y captura de los datos, pasando por una sucesión de ocupaciones similares a la incorporación, recolección y el filtrado de los mismos (pre-procesamiento), para luego ser procesados, analizados y evaluados hasta conseguir conocimiento agregada. Para eso, es requisito realizar un desarrollo iterativo que permita integrar varias consultas de selección (pre-procesamiento) de información, depuración de los datos, utilización de diferentes criterios de representación, además se aplican diferentes técnicas y algoritmos de MD, tanto descriptivas como predictivas.(Eckert & Suénaga, 2013)

6) Autor: Moris, Molen; Johan van Der

Título: Minería de datos educacionales: modelos de predicción del desempeño escolar en alumnos de enseñanza básica

Resumen:

En los años anteriores, se ha abierto una ocasión de llevar a cabo exámenes más precisos de las capacidades y desempeños de los alumnos. Poco a poco, comenzaron a proliferar tutores capaces y sistemas online que permitan registrar una cantidad enorme de información importante, referente al estudio de los estudiantes. La Minería de Datos Educacionales (MDE), es un campo de estudio destinado a desarrollar procedimientos matemáticos, para investigar datos que vienen de ambientes relacionados a la educación; y obtener la más grande proporción de información para intentar comprender mejor a los alumnos, instructores y actores relacionados, con el objetivo de hacer mejor los procesos académicos. Abordando el inconveniente de adivinar el desarrollo de un alumno dado sus datos históricos recopilados, desde su interacción en un sistema computacional de ejercitación online. Este propósito difícil de cumplir se construyó recientemente como posiblemente el más destacable dentro de la MDE, de la misma forma que prueba el aumento de las publicaciones similares, y el enorme interés que ha despertado de parte de las instituciones educativas y entidades gubernamentales. En esta proposición, se analizan los registros guardados de bastante más de medio millón de ejercicios online completados semanalmente en el 2011 por 805 alumnos en 23 tutoriales de cuarto básico de 13 academias atacables, explorando numerosos de los enfoques más usados para combatir este problema; y sugiriendo novedosas variedades para hacer mejor los resultados y contribuir a la descubrimiento de visualizaciones anómalas que podrían integrar instancias de "gaming the system". Complementariamente, se estudia el inconveniente de comprender cómo algunos contenidos chocan en otros. Hablamos de un inconveniente de Minería de Datos Educacionales central en el diseño curricular y la idealización de clases. Comúnmente esta red de influencias se crea basado en las críticas de profesionales. Algunos contribuyen explicitando la dependencia lógica de los contenidos, y otros con sus vivencias personales al enseñar esos contenidos. No obstante, es primordial contrastar esas críticas con el desarrollo de estudio que acertadamente suceden en el sala y crear redes causales basado en la prueba empírica. Al final, se reporta el examen del encontronazo de la ejercitación online en el desarrollo de la prueba SIMCE. Mediciones en condiciones de laboratorio detallan que la ejercitación aumenta el estudio. No

obstante, implementaciones institucionales no enseñaron impactos positivos. Este trabajo, exhibe la vivencia con academias atacables donde los alumnos hacen incontables ejercicios matemáticos por semana en un sistema online. El SIMCE de matemáticas subió de manera significativa, bastante más de tres ocasiones el incremento histórico logrado en todo el país en 2011. Además, los tutoriales que han realizado más grande proporción de ejercicios lograron un más grande encontronazo en el SIMCE, sin dependencia del efecto del docente y de la escuela. (Molen Moris, 2013)

7) Autor: Sanchez Merida, Cesar Julio

Título: Adaptación de estándares de dirección de proyectos particularizados para la minería de datos

Resumen:

Actualmente, el campo de la minería y el examen de datos está en apogeo. Cada vez más organizaciones apelan a la minería de datos para conseguir virtudes sobre sus competidores, circunstancia que se ve reflejada por el aumento de la estipulación laboral de personal especializado de parte de las compañías, independientemente de su sector de negocio, de esta forma como por el caudal de dinero movido por la industria enfocada. Este apogeo del área contrasta con el estancamiento de las metodología de administración de proyectos de minería de datos, las cuales bastante más de 20 años sin padecer cambios importantes, más allá de los cambios tecnológicos o los últimos movimientos del área, el cual, en los años anteriores, dejó de estar limitado a unos pocos expertos de las “Tecnologías de la Información”, siendo adoptado por compañías de los sectores más distintos, extendiéndose por todas las superficies departamentales de las organizaciones y por medio de los cambios tecnológicos; estas técnicas de administración de datos pasaron inclusive a tener aceptación por expertos no particulares del área. Sabiendo que todos los datos señalan que esta inclinación está en incremento se ve, por consiguiente, favorable comprobar la vigencia de las formas habituales de administración de proyectos de minería y examen de datos, con el objetivo de detectar sus debilidades y fortalezas, de esta forma como de comprobar su vigencia en el marco de hoy. Por medio del presente Trabajo de Final de Máster recorreremos las Formas de Dirección de Proyectos de Minería de Datos más empleadas, analizando sus debilidades, fortalezas y puntos de optimización, enumeraremos los intentos que se dieron de parte de los estudiosos con el objetivo de

ofrecer satisfacción a las inconsistencias y propondremos aquellas técnicas que suponemos tienen la posibilidad de ser más exitosas para ofrecer satisfacción a las mismas. Por último, propondremos una metodología que esperamos logre ofrecer satisfacción a las debilidades encontradas en la administración clásica de los proyectos de minería de datos. (Sánchez & César, 2017)

8) Autor: García-Martínez, R., Lelli, R., Merlino, H., Cornachia, L., Rodríguez, D., Pytel, P., Arboleya, H.

Título: Ingeniería de proyectos de explotación de información Para pymes

Resumen:

Los trabajos de explotación de información tienen propiedades muy dispares a las de los proyectos de avance de programa habituales. Las tradicionales etapas de examen, diseño, avance, incorporación y testeo, no tienen dentro con las etapas naturales de los procesos de avance de esta clase. De hecho, utilidades de la Ingeniería de Programa tradicional tales como: la ingeniería de requerimientos, los modelos de procesos, los ciclos de vida y también los mapas de ocupaciones no son ajustables a esta clase de proyectos. En este contexto, esta exploración facilita desarrollar y sistematizar el cuerpo de conocimiento de la Ingeniería de Proyectos de Explotación de Información con focalización en su transferencia a la Industria, especialmente al área PyMEs. Usando las formas de exploración documental exploratoria, prototipado evolutivo y casos de estudio se expone por medio de objetivos particulares, el avance de los próximos artefactos de Ingeniería de Proyectos de Explotación de Información: [a] una batería de técnicas de educación y formalismos de documentación de requerimientos; [b] un modelo de procesos y las métricas asociadas; [c] un modelo de ciclo de vida; y [d] un mapa de actividades. (García-Martínez, R., et al., 2011)

9) Autor: José Alberto Gallardo Arancibia

Título: Metodología para la definición de requisitos en proyectos de data mining

Resumen:

En años recientes, se dio inicio al avance de una cantidad sustancial de proyectos de Data Mining y distintos estudios estiman que esta cantidad, se incrementará más adelante inmediato. Las causas que impulsan este desarrollo son numerosas; principalmente, la considerable suma de datos que se desarrollan y guardan todos los días en las bases de datos de las organizaciones; la imposibilidad de procesar e investigar estos enormes volúmenes de datos por medio de procedimientos clásicos de análisis de datos y la necesidad de las compañías de conocer en ellos, patrones, relaciones, reglas o asociaciones servibles, que aporten información importante o conocimiento para el desarrollo de toma de decisiones. Cuando comienza un emprendimiento de Data Mining, la educación, el análisis y el modelado de los requisitos del usuario (proceso de Ingeniería de Requisitos), conforman ocupaciones importantes para el éxito de trabajos. No obstante, estas ocupaciones, comúnmente son las menos exploradas gracias a la inexistencia de técnicas, métodos o procedimientos ad-hoc para estos fines. Más allá de que es verdad, que en la labor “Evaluación de la Situación”, correspondiente a la primera etapa (Comprensión del Negocio) del modelo de desarrollo nivel CRISP-DM (Cross Industry Standard Process for Data Mining), uno de los procesos más ampliamente utilizados en los ámbitos industrial y académico, se propone como actividades iniciales, emitir un inventario de los recursos y establecer los requisitos, supuestos y restricciones del proyecto, no se menciona, cómo estas tareas deben ser desarrolladas, ni mediante qué instrumentos. En consecuencia, dada la importancia de contar con un documento eficaz de especificación de requisitos antes de dar inicio a un proyecto de Data Mining y la necesidad de disponer de un procedimiento metodológico explícito para obtener este documento, se ha desarrollado el presente trabajo de tesis doctoral, el cual propone una metodología para definir los requisitos de un proyecto de Data Mining. Esta metodología está centrada en las ocupaciones esenciales de la Ingeniería de Requisitos y se focaliza en la iniciativa de un Framework que radica de un desarrollo iterativo, conformado por un grupo de fases, que van desde la etapa de comprensión del dominio de negocio, <p>la de modelado del desarrollo decisional en la organización, hasta la etapa de creación del archivo final de requisitos. Para el avance de todas las fases del framework, se ofrecen un grupo de técnicas y artefactos. La metodología, fue ensayada en el avance de un emprendimiento

de Data Mining; su aplicación permitió detallar lazos de seguridad entre el cliente y programadores del emprendimiento, clarificar las ideas en relación al inconveniente y sus resoluciones, por medio de el modelado del desarrollo de negocios decisional, la posterior educación y validación de los requisitos, la observación por parte del cliente de resultados preliminares que permitieron la necesaria reabastecimiento y por último la redacción conjunta del archivo final de contrato, el cual fue suscrito de conformidad por el cliente y los programadores del emprendimiento. (Gallardo Arancibia, 2009)

10) Autor: Juan Miguel Moine

Título: Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo.

Resumen:

Para realizar en forma sistemática el desarrollo de hallazgo de conocimiento en bases de datos, popular como minería de datos, es necesaria la utilización de una metodología. En la actualidad las formas para minería de datos están en etapas tempranas de madurez, aunque algunas como CRISP-DM ya son usadas exitosamente por los equipos de trabajo para la administración de sus proyectos. En este trabajo se establece un exámen comparativo entre las formas de minería de datos más difundidas hoy en día. Para conseguir esa labor, y como aporte de esta proposición, se propuso un marco comparativo que explicita las propiedades que se deberían tomar en cuenta al instante de llevar a cabo esta confrontación. (Moine, 2013)

11) Autor: Rodríguez Montequín, M^a Teresa; Álvarez Cabal, J. Valeriano; Mesa Fernández, José Manuel; González Valdés, Adolfo

Título: Metodologías para la realización de proyectos de data mining.

Resumen:

La considerable suma de datos guardados en la actualidad en las organizaciones, unido al enorme avance tecnológico de las PCs, ha supuesto la aparición de novedosas opciones, agrupadas bajo el concepto por lo general popular como “data mining”. El aprovechamiento de estos datos necesita el avance de proyectos con propiedades particulares. Los proyectos de *Data Mining* tienen por propósito obtener información servible desde enormes proporciones de datos y se aplican a todos los sectores y en todos los campos. De esta forma hay proyectos de esta clase en sectores tan dispares como el comercio electrónico, la banca, las compañías industriales o la navegación

petrolífera. La extracción de esta información servible es un desarrollo complejo, que necesita la aplicación de una metodología estructurada para la utilización organizada y eficaz de las técnicas y utilidades accesibles.

En este texto se muestran las primordiales formas usadas por los investigadores para la ejecución de proyectos de *Data Mining*: CRISP-DM y SEMMA. Estas formas distribuyen la misma esencia estructurando el emprendimiento de Data Mining en fases que están interrelacionadas entre sí, convirtiendo el desarrollo de Data Mining en un desarrollo iterativo y amigable. La exhibición de las distintas fases y tareas de cada metodología brinda un concepto más extenso en relación a la ejecución de proyectos de *Data Mining*, que hará más fácil la adaptación de las metodologías, al avance de los proyectos de *Data Mining* particulares de cada organización. De esta forma, la exhibición de las fortalezas y debilidades de todas las formas hace viable la selección informada de una técnica de avance correcta para cada caso.(Montequín et al., s/f)

12) Autor: Juan Ángel Vanrell, Rodolfo Bertone, Ramón García-Martínez

Título: Modelo de proceso de operación para proyectos de explotación de información.

Resumen:

Los proyectos de explotación de información poseen características muy distintas a las de los proyectos de desarrollo de software tradicionales. Las clásicas etapas de análisis, diseño, desarrollo, integración y testeo no encajan con las etapas naturales de los procesos de desarrollo de este tipo de proyectos.

En este contexto, se propone un marco teórico para la creación de modelos de procesos de operación para proyectos de explotación de información para PYMEs siguiendo los lineamientos del modelo de procesos para la industria de software (COMPETISOFT).(Vanrell & Bertone, 2010)

13) Autor: Givanildo Santana do Nascimento; Adicinéia Aparecida de Oliveira

Título: AGILEKDD: An agile process model to knowledge discovery in databases and business intelligence systematization.

Resumen:

En el contexto de las economías basadas en el conocimiento y la sociedad del conocimiento, la competencia global se basa cada vez más en la capacidad de transformar datos en información, información en conocimiento y conocimiento en valor. Datos, información y conocimiento constituyen activos intangibles fundamentales para todas las organizaciones que trabajan en este modelo social y económico. En este contexto, la misión de Ingeniería de Software es producir sistemas capaces de procesar grandes volúmenes de datos, transformarlos en relevantes conocimiento y entregarlos a los clientes, para que puedan tomar decisiones correctas a la derecha hora. El desarrollo de este tipo de sistemas debe tener la orientación de un proceso capaz de conducir la transformación de los requisitos comerciales de los clientes en conocimiento explícito y productos de software, observando restricciones de tiempo, presupuesto y calidad más difíciles. El descubrimiento de conocimiento en bases de datos y el esfuerzo de sistematización de inteligencia de negocios ha dado como resultado varios modelos de proceso. Sin embargo, las empresas aún enfrentan fallas en la determinación del modelo de proceso utilizado en su descubrimiento de conocimiento en bases de datos y negocios. Los proyectos de inteligencia Los procesos disponibles aún no consideran Ingeniería de Software. Capacidades fundamentales como proyectos, requisitos y disciplinas de gestión de cambios. Varios procesos existentes no son adecuados para los entornos de negocios en constante cambio o falta de experimentación científica en casos reales, para confirmar sus cualidades y identificar sus deficiencias. El proceso propuesto en este trabajo, el AgileKDD, apunta a integrar las mejores prácticas de los principales procesos de descubrimiento de conocimiento en bases de datos con un proceso ágil de software. La aplicabilidad de AgileKDD fue verificada por un estudio de caso real, en que problemas comunes tales como los requisitos cambian y la mala calidad de los datos influyó en los resultados del proyecto. El caso de estudio señaló una mejora en el proceso. Las necesidades, que fueron consideradas en el refinamiento de AgileKDD. El proceso refinado resultante puede aplicarse como un marco adaptable y flexible para desarrollar sistemas de software capaces de descubrir conocimiento a partir de datos e información. El proceso respalda los principios y entrega continua de valor al cliente por medio de un proceso iterativo

e incremental ciclo de vida, respuesta inmediata a los cambios, así como la adaptabilidad y flexibilidad intrínseco a los procesos ágiles.(Nascimento & Oliveira, 2013)

14) Autor: Hernán Arboleya

Título: Propuesta de Ciclo de Vida y Mapa de Actividades para Proyectos de Explotación de Información.

Resumen:

La Explotación de la Información es una subdisciplina informática que le aporta herramientas a la Inteligencia del Negocio para poder transformar la información en conocimiento que sirva de base para llevar a cabo la toma de decisiones dentro de las organizaciones. Dado que este tipo de proyectos se diferencian a los proyectos convencionales de construcción de software y que las metodologías existentes para Proyectos de Explotación de Información no contemplan ni tiene en cuenta los riesgos que dentro de los mismos pueden ocurrir, en esta investigación se propone el diseño de un Modelo de Ciclo de Vida específico que tenga un análisis de los riesgos. Posteriormente se plantea la construcción de un Mapa de Actividades teniendo en cuenta el Modelo de Ciclo de Vida desarrollado y los procesos del Modelo de Procesos seleccionado. (Arboleya, 2013)

1.4. Justificación de la investigación

El método propuesto que ha sido elaborado en base a la Guía del PMBOK y la metodología CRISP-DM denominado MoPro-DM. Asimismo, este método permitió aportar conocimientos a las mejores prácticas de la gestión de proyecto de manera específica en la aplicación de proyectos de minería de datos en entornos educativos. De tal manera que contribuye a desarrollar un proyecto en dos subprocesos el subproceso de administración y el subproceso de Desarrollo. El modelo MoPro-DM propuesto podrá ser aplicado en el desarrollo de proyectos de minería de datos en cualquier universidad. De manera que se pueda lograr el cumplimiento de las actividades planificadas. Además, se tendrá un proceso bien detallado con las entradas y salidas en cada actividad del proyecto de minería de datos.

1.5.Limitaciones de la Investigación

Alcance

La investigación realizada se enfocará a los datos académicos de la Universidad Autónoma del Perú.

Limitación Tiempo

El estudio abarca los datos académicos desde los 2017-2018

1.6. Objetivos

1.6.1. Objetivo general

Determinar de qué manera el desarrollo de un modelo MoPro-DM mejora el proceso de aplicación de minería de datos académicos en las universidades peruanas en comparación con otros modelos

1.6.2. Objetivos específicos

- Determinar en qué medida el modelo mejora el Porcentaje de Cumplimiento Metodológico
- Establecer en qué medida el modelo mejora el Porcentaje de Cumplimiento de la Gestión de Proyecto
- Determinar en qué medida el modelo Porcentaje de cumplimiento en actividades de Desarrollo de MD

1.7.Hipótesis

1.7.1. Hipótesis general

El desarrollo de un modelo MoPro-DM mejora significativamente el proceso de aplicación de minería de datos académicos en las universidades peruanas en comparación con otros modelos.

1.7.2. Hipótesis específicas

- El modelo permite mejorar significativamente el % de cumplimiento Metodológico a nivel de Fases.
- El modelo permite mejorar significativamente el Porcentaje de cumplimiento de la gestión del proceso de minería de datos.
- El modelo permite mejorar significativamente de Porcentaje cumplimiento en actividades de Desarrollo de MD.

CAPITULO II:

MARCO TEÓRICO

2.1. Marco Conceptual

2.1.1. Teorías generales

2.1.1.1. Minería de Datos

La Minería de Datos se apoya en el concepto KDD - Knowledge Database Discovery, o Hallazgo de Conocimiento en Bases de Datos, KDD fue concebido en 1989 para resaltar que el saber es el resultado definitivo del hallazgo apoyado en los datos (Fayyad, 1996); según este creador la Minería de Datos se encuentra dentro de las etapas dentro del desarrollo KDD, en cambio la Minería de Datos se apoya en la aplicación de algoritmos para obtener patrones, KDD comprende todo el desarrollo de hallazgo de conocimiento, introduciendo todas sus fases: selección, pre procesamiento, transformación, Minería de Datos e interpretación / evaluación, de la misma forma que se expone en la Figura 4.

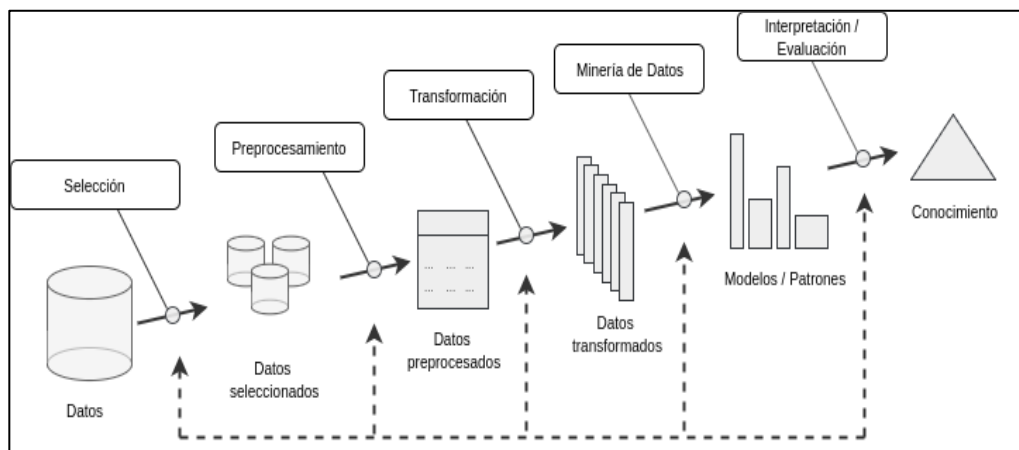


Figura 4. Pasos que componen el proceso KDD (Fayyad, 1996)

KDD facilita hallar patrones noticiosos y servibles dentro de un grupo de datos, estos patrones posibilitan el hallazgo para crear un conocimiento de todas formas válido y importante aplicado a un preciso problema; en ese sentido KDD se ocupa además de administrar puntos que van desde el alojamiento, extracción y rehabilitación de los datos, hasta la interpretación de los resultados de la Minería de Datos tras lo cual se crea el saber, todo este desarrollo se organiza en los próximos pasos:

Paso 1: Selección, se ajusta a la captura del grupo de datos sobre los que el hallazgo va a ser hecho.

Paso 2: Preprocesamiento, en cambio radica en hacer operaciones de depuración y limpieza sobre el grupo de datos con el objetivo de que sean firmes.

Paso 3: Transformación, radica en la ejecución de operaciones de reducción de dimensionalidad.

Paso 4: Minería de Datos, radica en la búsqueda de patrones sobre los datos, por medio del empleo de técnicas estadísticas, machine learning, etc.

Paso 5 y 6 : Interpretación / evaluación radica en la interpretación y evaluación de los patrones extraídos.

2.1.1.2. Analítica

Según (Campbell, DeBlois, & Oblinger, 2007) exponen que la analítica se relaciona con grandes conjuntos de datos, técnicas estadísticas y modelos de predicción, de forma se puede pensar como la práctica de la minería de datos institucionales que produce "inteligencia procesable". Hoy en día el término "analítica" se aplica de forma frecuente a los esfuerzos en la explotación de datos de diversas fuentes para ayudar a las instituciones a ser más eficientes. Así, cada vez diferentes industrias están desarrollando esta estrategia para lograrlo (Thomas H Davenport, Harris, & Morison, 2010). (Bichsel, 2012), además de colocar el énfasis en la predicción y siguiendo esta idea, define la analítica como:

“El uso de los datos, el análisis estadístico y los modelos explicativos y predictivos para obtener información y actuar sobre temas complejos” (Bichsel, 2012).

Con mayor frecuencia se relaciona la analítica como "estratégica" y esta participa en la "toma de decisiones". (Bichsel, 2012), menciona que algunos sugirieren que se deben mencionar los tipos de datos e informes que participan con la analítica, y sugiere que la "visualización de datos" debe ser parte de su definición. Se afirma que la analítica es un proceso que contiene un número de pasos a seguir; de esta forma (Bichsel, 2012), describe ese proceso como:

- 1) Comenzar con una cuestión/pregunta estratégica.
- 2) Encontrar o recolectar los datos adecuados para responder a dicha pregunta.
- 3) Analizar los datos con miras a la predicción y al descubrimiento de nuevo conocimiento.
- 4) Representar los resultados de manera que sean comprensibles y aplicables.
- 5) Realimentar el proceso para abordar la pregunta y cuestiones estratégicas y para crear otras nuevas.

La analítica puede ser fuente de información para la toma de decisiones humanas o puede soportar la toma de decisiones de forma automatizada.

La analítica es subconjunto de la inteligencia de negocios (del inglés, *Business Intelligence*, BI), un conjunto de tecnologías y procesos que utilizan datos para comprender y analizar el rendimiento del negocio (Thomas H Davenport, Harris, & Shapiro, 2010).

La necesidad de utilizar analítica ha surgido en el momento en que las empresas han ofrecido productos similares y han usado tecnología comparable, de manera que los procesos de negocio de alto rendimiento se convierten en algunos de los últimos puntos en los que se puede conseguir una diferenciación. Muchas de las bases previas para la competencia ya no están disponibles, como la ventaja geográfica única que ya no importa en la competencia mundial, y la regulación proactiva ha desaparecido en gran medida.

Las tecnologías patentadas se copian rápidamente, y romper la innovación en productos o servicios parece cada vez más difícil de lograr. Lo que queda como una base para la competencia consiste en ejecutar su negocio con la máxima eficiencia y eficacia, así como tomar las decisiones de negocio lo más inteligente posible. Entonces la analítica en la competencia se ejerce al exprimir hasta la última gota de valor a partir de los procesos de negocio y las decisiones clave y se define como una organización que utiliza la analítica amplia y sistemáticamente para ejercer la competencia (Thomas H Davenport, Harris, & Shapiro, 2010).

Desde este punto de vista empresarial, Thomas H. Davenport (2006) menciona que las compañías y organizaciones están sacando partido a sus bases de datos. Tras acumular dichos datos durante años y darse cuenta de que necesitaban mejorar su capacidad transaccional, muchas organizaciones han llegado a un punto en el que han pensado (T. H Davenport, 2006):

De acuerdo, no tenemos más excusas; tenemos que empezar a gestionar nuestro negocio de una manera diferente, basándonos en nuestros datos. La segunda conclusión es que toda compañía puede mejorar sus capacidades analíticas. El análisis avanzado puede llegar a ser una parte integral de la estrategia de negocio. El tercer elemento que se necesita además de datos y tecnología para llegar a competir en análisis avanzado, y que son los factores diferenciales, tienden a ser humanos, como por ejemplo: el liderazgo, la

pasión, las habilidades y las relaciones; aspectos que, creo, han sido desatendidos en las investigaciones sobre inteligencia de negocios, también llamada analítica, de los últimos años. Las compañías que cuentan con una alta sofisticación analítica suelen contar con empleados altamente cualificados trabajando para ellas. Esto no creo que sea casual (T. H Davenport, 2006).

El potencial de la analítica de acuerdo con estas definiciones es que ayuda al evaluar las acciones pasadas para estimar el potencial de las acciones futuras, con las cuales tomar mejores decisiones y adoptar estrategias más eficaces, ya sea a nivel organizacional o individual.

Un factor menos popularizado a la hora de realizar la explotación efectiva de la analítica es la rica variedad y madurez de las técnicas de análisis de datos. Un analista experto tiene ahora muchas disciplinas en las que inspirarse y muchas herramientas al alcance para usarlas.

La diversidad y la flexibilidad de algunas de las técnicas que se alinearon bajo la bandera analítica se evidencia por las numerosas aplicaciones: los mercados financieros, análisis de los deportes, econometría, precios de los productos y la maximización del rendimiento, de fraude, de detección de la delincuencia, filtros de correo electrónico de (*spam*), comercialización, segmentación de clientes, la eficiencia organizacional. Detrás de estas aplicaciones se pueden encontrar las raíces de la analítica en el nacimiento de las estadísticas en el siglo XVIII, pero desde entonces las diferentes aplicaciones de la estadística y de las ICT han dado lugar a diferentes comunidades de práctica que ahora parecen estar fusionándose (Cooper, 2012).

Tabla 1
Preguntas cuyas respuestas son fuentes de información y conocimiento, mediante un análisis direccional

	Pasado	Presente	Futuro
Conocimiento/Información	¿Qué sucedió? Reporte y Descripción	¿Qué está sucediendo ahora? Alerta	¿Qué sucederá? Extrapolación
	¿Cómo y por qué sucedió? Modelos y Explicación	¿Cuál es la mejor acción a seguir? Recomendaciones	¿Qué es lo mejor o peor que puede suceder? Predicción

Fuente: (Thomas H. Davenport et al., 2010)

Para organizar los actuales y diversos enfoques que existen de la analítica, a pesar de que las preguntas clave que la analítica puede abordar son innumerables, se han definido una serie de preguntas clave para esquematizar los enfoques existentes. Estas preguntas esquematizan los diferentes enfoques de acuerdo a la dimensión temporal de las preguntas, así como su nivel de aportación de conocimiento. Estos aspectos y las preguntas claves se resumen en la matriz propuesta en (Thomas H Davenport, Harris, & Shapiro, 2010), que se recoge en la Tabla 2, que distingue entre enfoques más basados en la realidad (fila de la *información*, del inglés "*information*") y enfoques que están más orientados hacia la comprensión más profunda (la fila *conocimiento*, del inglés "*insight*") y los segmentos de estos de acuerdo a un marco temporal.

Los campos que algunos ven anticuados, como por ejemplo la investigación operativa, y otros que a menudo se perciben tan futuristas como inteligencia artificial, están haciendo contribuciones de formas sorprendentes. Mientras tanto, la comunidad educativa ha hecho sus propias contribuciones; el análisis de redes sociales e inteligencia artificial, ambos han surgido de la investigación académica y ahora se está empezando a ver las variantes específicas del sector de la analítica que están siendo utilizados bajo la forma de la EDM, analítica del aprendizaje y bibliometría.

2.1.2. Bases teóricas especializadas sobre el tema

2.1.2.1. Minería de Datos Educativa (EDM) y Analítica de Aprendizaje (LA)

EDM (Educational Data Mining) ha sido una disciplina en evolución, desde 2009 ha contado con una comunidad de investigadores que ha crecido y se ha desarrollado, esto ha permitido incluso que desde ese mismo año se lleve a cabo una serie de eventos anuales denominadas EDM, que constituyen también su propio Journal JEDM (Journal of Educational Data Mining), después la publicación del libro Handbook of Educational Data Mining de Chapman & Hall/CRC, creo otro hito importante que ha dado impulso a esta disciplina (R. S. J. d Baker & Yacef, 2009).

En cambio, la *Analítica de Aprendizaje* se ha definido como un área de investigación y aplicación que se relaciona con el análisis académico, el análisis de acciones y el análisis de predicción. La analítica del aprendizaje se basa en una gama más amplia de disciplinas académicas en comparación con la Minería de Datos Educativa; incorpora conceptos y técnicas de la ciencia de la información, la informática, la estadística, la psicología y las ciencias del aprendizaje.

El EDM corresponde a la aplicación de técnicas propias de Minería de Datos hacia el ámbito académico; es un paradigma orientado a la generalización de modelos, métodos y algoritmos para la exploración de datos propios de un contexto educativo, con el fin de encontrar y analizar patrones que definen a los estudiantes en base a sus logros, evaluaciones y el dominio de conocimiento alcanzado con los diversos mecanismos de aprendizaje - enseñanza que hoy en día están disponibles en la mayoría de instituciones educativas, sobre todo de nivel universitario. La Minería de Datos Educativa aporta a la generación de modelos educativos en los cuales se fomenten nuevas técnicas y herramientas que mejoren el nivel participativo de los estudiantes, (Ballesteros, Alejandro, Sánchez, Daniel, & García, Ricardo, 2013).

Lo que se destaca en los métodos de EDM, es que a menudo difieren de los estándares aplicados en Minería de Datos general, esto debido a que en EDM existe la necesidad de presentar explícitamente la *jerarquía multinivel* y la *dependencia* entre los datos académicos. (R. S. J. Baker, 2011).

2.1.2.2. Métodos de EDM

Las formas utilizadas en EDM son de fuentes distintas, ambas más destacables son por un lado *Minería de Datos* y *Analítica*, y por otro lado de *Psicometría* y *Medición*

Educativa. En numerosos casos, las propiedades particulares de los datos académicos dieron lugar a diferentes procedimientos que han desempeñado un papel más importante en EDM que en la Minería de Datos generalmente, o dieron lugar a adaptaciones a los procedimientos psicométricos que ya están. Ahora, se señala varios de los primordiales procedimientos aplicados en EDM. (Moscoso-Zea, Andres-Sampedro, & Luján-Mora, 2016).

Métodos de Predicción: El propósito es desarrollar un modelo que permita inferir puntos sencillos de los datos (variable predicha), basado en la conjunción de otras características de los datos (variables predictoras). Para desarrollar un modelo de predicción es requisito comprender los valores de la variable a adivinar para un grupo de datos reducido, llamado grupo de entrenamiento, el modelo entonces es generado para dicho grupo, después es validado con otro grupo de datos, el llamado grupo de prueba, con esa validación; y basado en la exactitud alcanzada, se va a poder determinar si el modelo puede pasar a producción. Los tres modelos de predicción frecuentes en EDM son: clasificadores (classifiers), regresores (regressors) y estimación de conocimiento latente (latent knowledge estimation). En clasificadores la variable predicha puede ser binaria o categórica, algunos procedimientos populares son: árboles de elección, bosques al azar, reglas de elección, regresión gradual y regresión logística. En regresión, la variable predicha es una variable continua (numérica), el regresor más habitual en EDM es regresión lineal. En estimación del conocimiento latente, el cual verdaderamente se ajusta a un tipo particular de clasificador, se analiza el saber de un estudiante sobre determinados puntos y conceptos particulares, basado en sus patrones de correctitud en esas capacidades. Los modelos usados en el estudio online acostumbran diferir de los modelos psicométricos usados en las pruebas de papel o en las pruebas adaptativas por PC, porque con una aplicación de estudio amigable, el saber del estudiante está modificando siempre.

Descubrimiento de estructuras: el propósito es conocer construcciones dentro de los datos sin tener un concepto a priori de que es lo que se va a encontrar, algo muy distinta al propósito en predicción, donde había una variable específica que el investigador trata de modelar, en cambio en hallazgo de construcciones no hay una

variable de interés específica, hablamos de saber qué composición se obtiene de manera natural de los datos. Los enfoques frecuentes son los siguientes: clustering, análisis de componentes, análisis de comunidades y hallazgo de construcciones de dominio. En clustering el propósito es hallar “puntos” en los datos que los agrupen naturalmente, formando los catalogados” clusters”, esta clase de algoritmos son servibles en casos donde las categorías frecuentes dentro de los datos no son populares antes, como entre otras cosas hallar agrupación de alumnos basado en diferentes criterios (Beal, 2006). Los modelos para análisis de comunidades (SNA) son desarrollados desde la relación e interacción entre los actores particulares, de esta forma como de los patrones que emergen de su relación e interacción. Entre otras cosas, para detectar proyectos grupales efectivos o no, se podría recurrir a investigar visualmente la fuerza de las conexiones del grupo. El hallazgo de construcciones de dominio radica en hallar la composición de conocimiento dentro de un dominio educativo. Entre otras cosas, como contenidos particulares tienen la posibilidad de mapearse a elementos particulares de conocimiento o capacidades, en los alumnos.(R. S. Baker, 2014).

Minería de relaciones: el objetivo es descubrir relaciones entre variables dentro de un conjunto de datos que posee un gran número de variables; se podría aplicar para identificar cuáles son las variables que poseen la relación más fuerte con una determinada variable de interés, o se podría buscar descubrir cuál relación entre dos variables es la más fuerte. Existen cuatro tipos de minería de relaciones: reglas de asociación, minería de correlaciones, minería secuencial de patrones, minería causal de datos (Moscoso-Zea, Andres-Sampedro, & Luján-Mora, 2016). En reglas de asociación el objetivo es encontrar reglas del tipo SI - ENTONCES, de modo que si cierto conjunto de valores en determinadas variables es encontrado, otra variables generalmente tendrá un valor específico. En minería de correlación el objetivo es encontrar correlaciones lineales positivas o negativas entre variables. En minería secuencial de patrones el objetivo es encontrar asociaciones temporales entre eventos. En minería causal el objetivo es encontrar si un evento es la causa de otro evento, por ejemplo, encontrar que factores afectan un pobre rendimiento de los estudiantes en la clase (Fancsali, 2012).

Descubrimiento con modelos: el modelo construido en base a un análisis de Minería de Datos, es aplicado sobre datos para evaluar el fenómeno que el modelo identifica,

la predicción entonces es utilizada como entrada en otro método de Minería de Datos. Por ejemplo, utilizar los resultados de un modelo predictivo como variable predictora en un segundo modelo de predicción y esto en múltiples niveles (R. S. Baker, 2014).

2.1.2.3. Minería de datos para la educación

Al igual que con los primeros esfuerzos para entender los comportamientos en línea, los primeros intentos de minería de datos de datos educativos involucraban la minería de los web log (R. S. J. d Baker & Yacef, 2009), pero ahora más integrados, instrumentados y sofisticados, los sistemas de aprendizaje en línea ofrecen más tipos de datos. Una característica importante y única de los datos sobre la educación es que son jerárquicos. Los datos a nivel de pulsación (del inglés, click), nivel de respuesta, nivel de sesión, nivel de los estudiantes, nivel de la clase y el nivel de los profesores y los centros educativos están incluidos y relacionados unos dentro de otros (R. S. J. d Baker & Yacef, 2009). Otras características importantes son el tiempo, la secuencia y el contexto.

El tiempo es importante para capturar datos, tales como la duración de las sesiones de práctica o tiempo para aprender. La secuencia representa cómo los conceptos se construyen unos detrás de otros y cómo las prácticas y tutorías deben ser ordenadas. El contexto es importante para explicar los resultados y saber que un modelo puede o no puede funcionar.

Los métodos de minería de datos jerárquica y longitudinal de modelado de datos han sido avances importantes en la EDM. La minería de datos para la Educación en general hace hincapié en la reducción del aprendizaje en pequeños componentes que pueden ser analizados y luego influidos por el software que se adapta al estudiante (Moscoso-Zea et al., 2016). Los datos del aprendizaje de los estudiantes recogidos por los sistemas de aprendizaje en línea se están explorando para desarrollar modelos predictivos mediante la aplicación de métodos de minería de datos educativos que clasifican los datos o encuentran relaciones. Estos modelos tienen un papel clave en la construcción de sistemas de aprendizaje adaptativos, en los que las adaptaciones o las intervenciones basadas en las predicciones del modelo se puede utilizar para cambiar lo que los estudiantes experimentan o incluso recomendar experiencias fuera de los servicios académicos para apoyar su aprendizaje (García, Enrique, Romero, Cristóbal, De Castro Lozano, Carlos, Calders, Toon, & Ventura, Sebastian, 2011).

La EDM se está convirtiendo en un área de investigación con un conjunto de métodos computacionales y psicológicos además de diversos planteamientos de investigación para entender cómo los estudiantes aprenden. Nuevos métodos interactivos de aprendizaje basados en ordenador y herramientas (sistemas inteligentes de tutoría, simulaciones, juegos) han abierto oportunidades para recopilar y analizar datos de los estudiantes, para descubrir patrones y tendencias en los datos, y para hacer nuevos descubrimientos y probar hipótesis sobre cómo los estudiantes aprenden. Los datos recogidos de los sistemas de aprendizaje en línea se pueden agregar más de un gran número de estudiantes y pueden contener muchas variables que los algoritmos de minería de datos pueden explorar para construir modelos.

Un significado, más restrictivo para la minería de datos educativa, que se refiere al uso de técnicas de minería de datos es el de (Romero & Ventura, 2007): para analizar datos con el fin de resolver los problemas de investigación educativa. La EDM se refiere al desarrollo de métodos para explorar los tipos de datos de los centros educativos y, con estos métodos, comprender mejor a los estudiantes y las características del entorno en el que aprenden (Romero & Ventura, 2007). La EDM explota la estadística, las máquinas de aprendizaje y los algoritmos de minería de datos sobre los distintos tipos de datos educativos. El Departamento de la Tecnología Educativa de EE.UU. ofrece una discusión de minería de datos educativos y, en general, sobre la analítica del aprendizaje (Marie Bienkowski, Mingyu Feng, & Barbara Means, 2012)): *EDM desarrolla métodos y aplica las técnicas de la estadística, aprendizaje automático y minería de datos para analizar los datos recogidos durante la enseñanza y el aprendizaje. EDM pone a prueba las teorías de aprendizaje e informa a la práctica educativa. LA aplica técnicas de la informática, la sociología, la psicología, estadística, aprendizaje automático y minería de datos para analizar los datos recopilados durante los servicios educativos y la administración de la educación, la docencia y el aprendizaje. LA crea aplicaciones que influyen directamente en la práctica educativa.*

El desarrollo de la analítica del aprendizaje se ha visto impulsado al incluirlo en el informe Horizon del 2012 del NMC (Johnson, L., Adams, S, & Cummins, M, 2012)

Este informe es parte de una serie y el NMC es una empresa establecida en el 2002 dedicada a identificar y describir las tecnologías emergentes y su potencial impacto en los próximos cinco años en una variedad de sectores en todo el

mundo. Específicamente, el reporte Horizon del 2012 examina las nuevas tecnologías aplicadas a la educación, el aprendizaje y en la investigación creativa y su potencial impacto en estas áreas.

2.1.2.4. Modelo de referencia para la Analítica del Aprendizaje

El marco de referencia tiene la intención de ser una guía y un descriptor de las zonas problemáticas. Por tanto, puede referenciarse como un "marco de diseño" que puede y debe utilizarse para el diseño de servicios de LA desde una perspectiva integradora. En este mismo sentido Chatti et al. (2012) describen un modelo de referencia para LA basado en cuatro dimensiones que identifica diversos retos y oportunidades de investigación en el área de LA en relación con cada dimensión. Como se representa en la Figura 7, las cuatro dimensiones del modelo de referencia propuesto para LA son:

- ¿Qué?: ¿Qué tipo de datos reúne el sistema para su análisis?
- ¿Quién?: ¿Quién participa en el proceso análisis?
- ¿Por qué?: ¿Por qué se requiere el sistema de análisis de los datos recogidos?
- ¿Cómo?: ¿Cómo funciona el sistema para realizar el análisis de los datos recogidos?

Algunos usos potenciales de las aplicaciones de LA proporcionan apoyo a:

- Identificar a los estudiantes en riesgo a fin de proporcionar intervenciones positivas diseñadas para mejorar la retención.
- Brindar recomendaciones a los estudiantes en relación con el material de lectura y actividades de aprendizaje.
- Detectar la necesidad de mejoras pedagógicas y medir los resultados de estas.
- Adaptar la oferta de cursos.
- Identificar los profesores que tienen un buen desempeño y los profesores que necesitan ayuda con los métodos de enseñanza.
- Ayudar en el proceso de reclutamiento de estudiantes.

El informe *Horizon* del *New Media Consortium* (Johnson et al., 2011) ofrece una definición más completa: *Analítica del aprendizaje se refiere a la interpretación de una amplia gama de datos producidos por y recopilados a favor de los estudiantes con el fin de evaluar su progreso académico, predecir el rendimiento futuro y los posibles problemas puntuales. Los datos se obtuvieron de las acciones explícitas*

estudiantiles, tales como completar las tareas y tomar los exámenes, de las acciones tácitas, que incluyen las interacciones sociales, actividades extracurriculares, mensajes en foros de discusión y otras actividades que no son evaluadas directamente como parte del progreso educativo del estudiante" (Johnson et al., 2011).

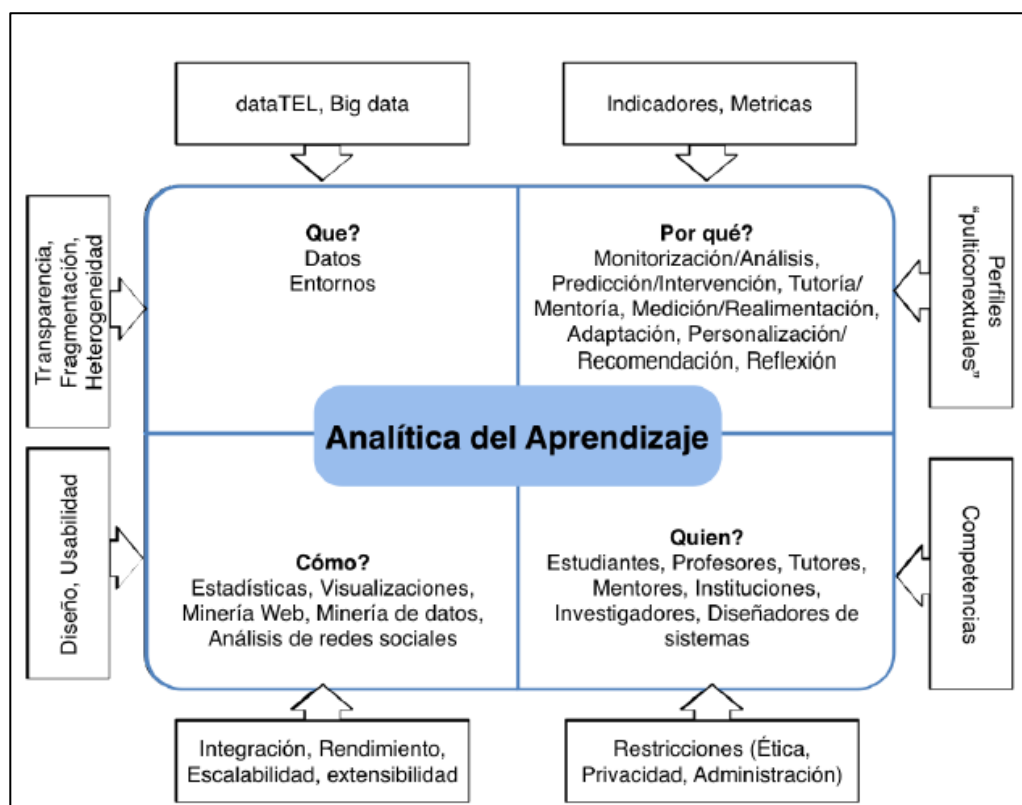


Figura 5. Modelo de referencia para la analítica del aprendizaje, modificado de (Chatti et al., 2012).

El objetivo de la analítica del aprendizaje es permitir que los profesores y los centros educativos adapten las oportunidades educativas al nivel de las necesidades y capacidades de cada estudiante de la manera más rápida posible. La promesa de la analítica del aprendizaje es la de aprovechar el poder de los avances en la minería de datos, la interpretación y modelado para mejorar la comprensión de la enseñanza y el aprendizaje, y para adaptar la educación individualmente a los estudiantes con una mayor eficacia.

2.1.2.5. Metodologías para análisis de datos

CRISP-DM

La metodología CRISP-DM analiza el proceso de explotación de la información en seis fases diferentes (Figura 6). Aunque en la ilustración se muestran las interacciones más comunes entre las fases se pueden establecer relaciones entre cualquiera de ellas (IBM, 2012)

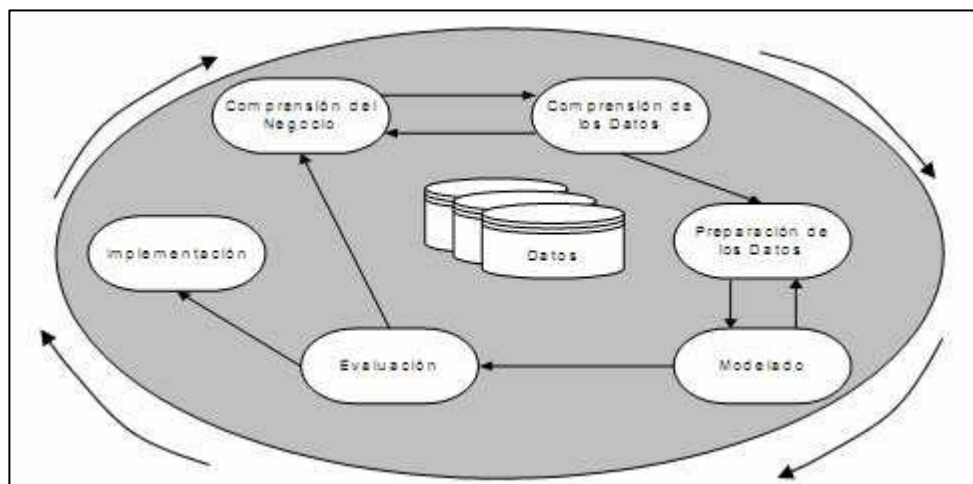


Figura 6. Metodologías para análisis de datos adaptado de (IBM, 2012)

SEMMA

La primera metodología fue propuesta por el SAS Institute, su nombre hace referencia a las cinco fases que se consideran al utilizarla (Sample, Explore, Modif., Model, Assess esto es Muestrear, Explorar, modificar, Modelar y Valorar respectivamente) (SAS Institute, 2018).

La figura 7 muestra la dinámica del método SEMMA.

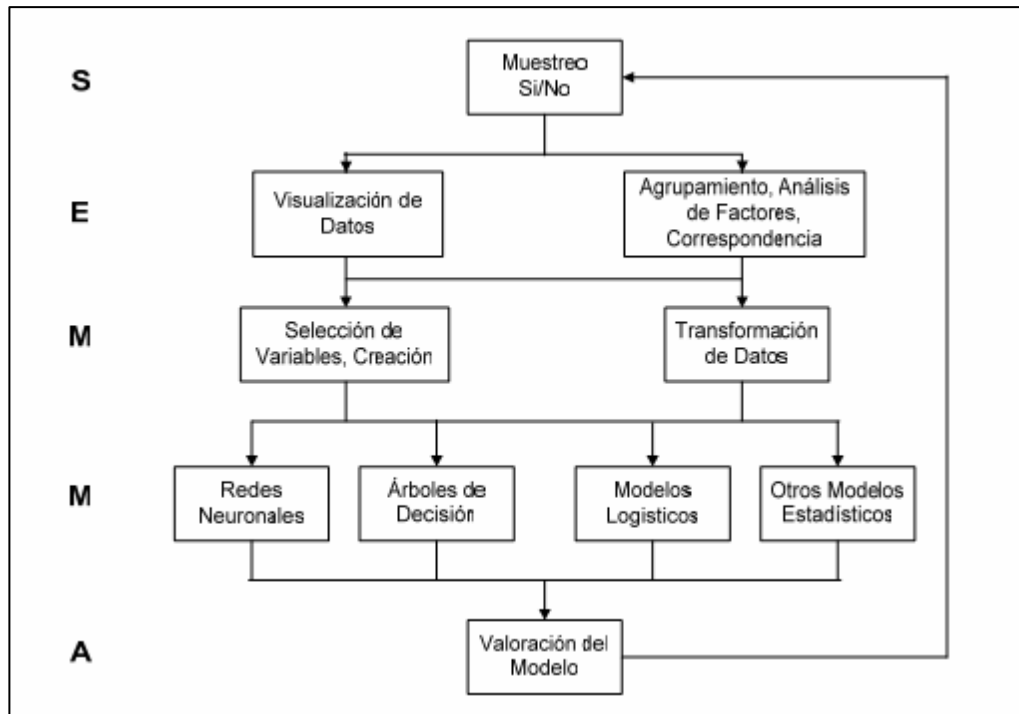


Figura 7. Semma.adaptado de (SAS Institute, 2018)

P3QT

La metodología de D. Pyle se divide en dos etapas, la primera denominada Modelado de Negocio o MII, y la segunda llamada Minería de datos o MIII (Pyle Dorian, 2003).

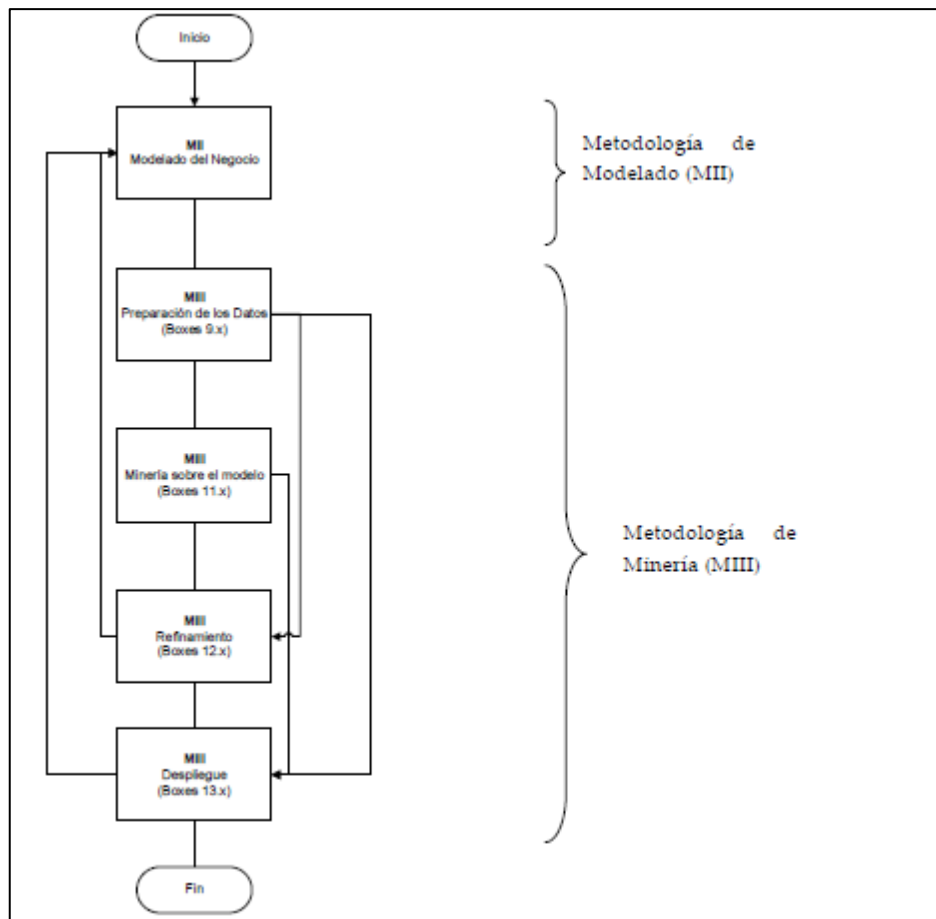


Figura 8. P3QT

Para comenzar la primera etapa Pyle propone cinco posibles puntos de partida en función del propósito del proyecto de explotación de la información que se quiere evaluar. De esta manera Pyle considera:

1. Explorar los datos en búsqueda de relaciones útiles.
2. Dada una oportunidad o problema ver cómo puede la explotación de la información encausar a la organización hacia una decisión correcta.
3. Simplemente ver qué puede lograr la explotación de la información
4. Utilizar la minería de datos para construir un modelo sobre una situación particular
5. Dada una situación estratégica, analizar si la minería de datos puede ser útil para explicar la situación y cuáles son las opciones de la organización para resolverla.

ASD - DM

Es un framework llamado ASD - DM propuesto en (Alnoukari & El Sheikh, 2012), este ofrece la conjunción de las propiedades propias del Avance de Programa Adaptativo o ASD, con los pasos en general usados en proyectos de Minería de Datos. La iniciativa de los autores radica en la estructuración de las etapas especificadas por CRISP-DM en las tres fases propias de ASD: Especulación, Colaboración y Estudio, como se expone en la figura 9. Estas fases aceptan que la administración del emprendimiento posea los próximos fundamentos: orientación a las propiedades, iteratividad, administración de peligros y tolerancia a los cambios.

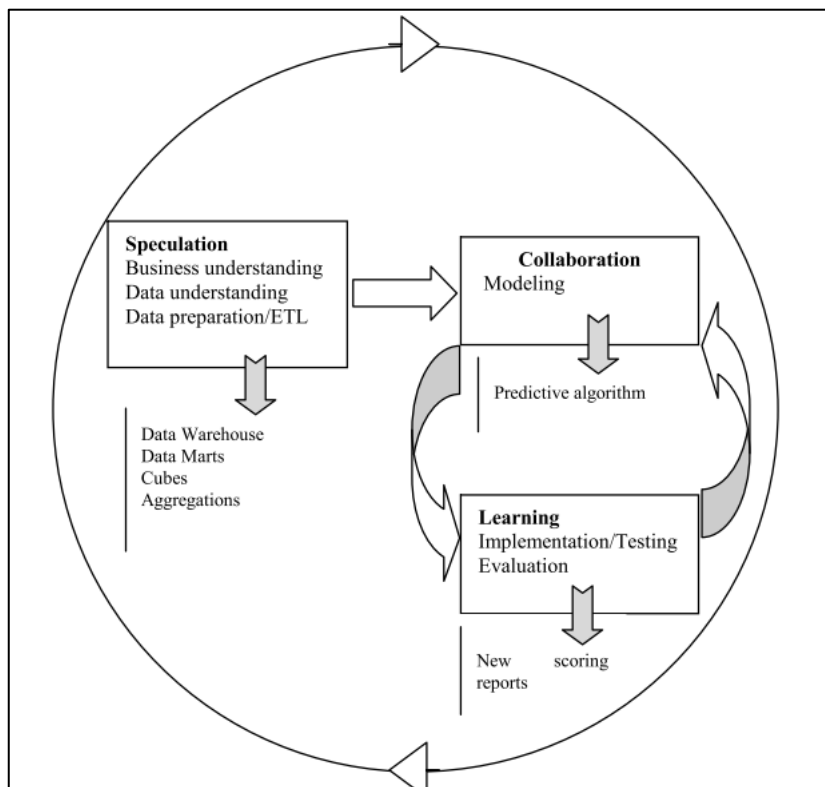


Figura 9. Framework para Minería de Datos Predictiva basado en el desarrollo de software adaptativo ASD

Lo atrayente de esta iniciativa es el ingrediente ágil que se busca insertar en lugar del período clásico Planifica - Diseña - Crea, que puede sospechar CRISP-DM. Y que tiene mucha coherencia ya que comúnmente un emprendimiento de Minería de Datos, tiene un aspecto considerable de indecisión sobre los datos accesibles, su calidad, la exactitud de los modelos, etc., puntos que únicamente con la ejecución del ensayo (Modelamiento) se podrían comprobar. Frente esta visión la etapa de idealización acaba siendo menos servible mientras más indecisión permanezca en el emprendimiento. seguramente lo verdaderamente acertado sería, una vez definido el inconveniente o pregunta, hacer un desarrollo

liviano sobre las primeras fases del desarrollo (comprensión del negocio), y llegar de manera rápida hasta la experimentación (comprensión de los datos y modelamiento), esto permitiría al analista meterse en el contexto del emprendimiento, los elementos accesibles y los resultados probables. Después de lo cual realizar una exclusiva iteración sobre todas las fases del desarrollo, y en ella determinar los puntos antes pasados por prominente o poco desarrollados, como sería caso de la idealización del emprendimiento.

Modelo de Proceso KDDA en caracol

Una iniciativa semejante a lo planteado en el final del punto previo es la abordada por Li Yan en su proposición doctoral “New Artifacts For The Knowledge Discovery Via Data Analytics (Kdda) Process”. El creador en el capítulo 4 muestra un modelo llamado “Modelo de Desarrollo KDDA en caracol” (A Snail Shell KDDA Process Model) concordante con las fases y tareas de CRISP-DM, por otro lado sus dos primordiales aportes corresponden a la marcada consideración que se le otorga a la primera etapa del emprendimiento Formulación del Inconveniente, y más adelante al carácter iterativo de las siguientes fases, como se puede ver en la Figura 10, ahí se hace explícita la oportunidad de que luego de cada etapa se puede de nuevo regresar a las fases iniciales para corregir o determinar algo en el desarrollo. Merece señalar que el mismo CRISP-DM considera de todas formas esta necesidad, y deja abierta la oportunidad de desplazarse entre las diferentes fases y ocupaciones basado en las particularidades de cada emprendimiento.

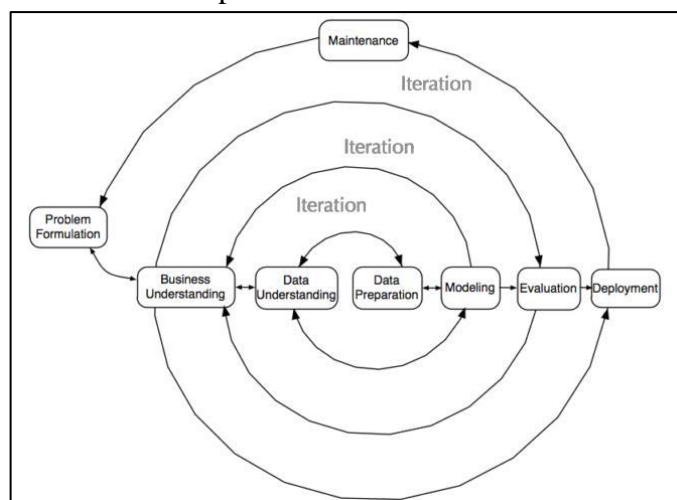


Figura 10. Modelo de Proceso KDDA en caracol

2.1.2.6. Clasificación de técnicas de minería de datos utilizadas en entornos educativos

Son múltiples los casos de estudio en los que se aplican técnicas de minería de datos en entornos educativos, cada uno de ellos buscando una respuesta a una situación particular que aporte al objetivo buscado por la EDM. Para el desarrollo del presente trabajo se tuvieron en cuenta trabajos publicados sobre EDM entre el año 1993 y el año 2015, de los trabajos revisados se tuvieron en cuenta 130 trabajos en los que se aplicaron al menos una técnica de minería de datos con el fin de analizar una situación asociada al proceso de enseñanza-aprendizaje. La revisión de literatura adelantada se organizó en cuatro apartados así:

1. Revisión de trabajos en los cuales se han utilizado técnicas de minerías de datos en entornos educativos.
2. Trabajos clasificados por dominios educativos.
3. Trabajos clasificados por técnicas de minerías de datos
4. Listado de autores relevantes en EDM.

2.1.2.7. Trabajos clasificados por técnicas de minerías de datos

En los trabajos analizados se encontró que en total se utilizan 13 técnicas de minerías de datos, estas son:

- Análisis de Correlación
- Árboles de decisión
- Árboles de Regresión
- Cadenas de Markov
- Clasificación
- Clustering
- Differential Sequence Mining
- Patrones Secuenciales
- Redes Bayesianas
- Redes Neuronales
- Reglas de asociación
- Regresión Lineal

Los trabajos analizados en la revisión de literatura están asociados a las diferentes técnicas de minería de datos, en algunos casos el mismo paper utilizó diferentes técnicas, el paper es

clasificado en tantas técnicas como haya utilizado para considerar todas las perspectivas utilizadas por los diferentes actores. En la tabla No. 3 se presentan todos los trabajos asociados a las técnicas enunciadas anteriormente.

Tabla 2
Trabajos clasificados por técnicas de minería de datos

Técnica Implementada	No. de Publicaciones	Autores Relevantes	Años con publicaciones
Análisis de Correlación	2	Chen C., Chen M., Patarapichayatham, C. Kamata, A. Kanjanawasee, S.	2009, 2012
Arboles de decisión	27	Baradwaj B., Pal S., Sheard J., Ceddia J., Hurst J., Tuovinen J., Guo Q., Zhang M., Kovačić, Z. J	Entre el 2003 y 2015
Arboles de Regresión	1	Hu Y., Lo C., Shih S.	2014
Cadenas de Markov	1	Stamper J., Barnes T.	2009
Clasificación	15	Bhardwaj B., Pal S., Marquez-Vera C., Romero C., Ventura, S., Chen C., Chen M., Bayer J., Bydzovská H.,	Entre el 2009 y 2015
Clustering	27	Wang F.H., Shao H.M., Baradwaj B., Pal S., Kumar V., Chadha A., Chen C., Chen M.	Entre el 2003 y 2015
Differential Sequence Mining	1	Martinez-Maldonado R., Yacef K. Kay J.	2013
Patrones Secuenciales	17	Karampiperis P., Sampson D., Romero C., Ventura S., Zafra A., De bra P., Kay J., Maisonneuve N., Yacef K., Zaiane O.R.	Entre el 2004 y 2015
Redes Bayesianas	10	Pardos Z. A., Heffernan N. T., Anderson B., Heffernan C. L., Schools W., Pandey U., Pal, S.	Entre el 2006 y 2013
Redes Neuronales	10	Oladokun V.O., Adebajo A.T., Charles-owaba O.E. , Hämäläinen W, Vinni M	Entre el 1993 y 2015
Reglas de asociación	35	Ha S., Bae S., Park S., Zaiane O., Romero C., Ventura S., De Bra P., Merceron A., Yacef K., Baradwaj B., Pal S.	Entre el 1995 y 2014
Regresión Lineal	2	Hämäläinen W, Vinni M	2010

Fuente: Elaboración propia

2.1.2.8. Clasificación de trabajos por dominios educativos.

Cada uno de los trabajos revisados fue clasificado en uno de los 7 dominios de tipo educativo que se tomaron para el desarrollo de la tesis. Es importante precisar que estos dominios surgen a partir de la revisión elaborada, cada trabajo revisado se asoció a un dominio de la educación. Los dominios de clasificación se seleccionaron a partir de los temas que tratan los trabajos, estos dominios son:

- **Análisis de Deserción o Retención:** trabajos en los cuales se analizan causantes relacionados a deserción académico o estudios que buscan beneficiar la retención académica.
- **Análisis de OVA o EVA:** esos trabajos en los cuales se analizaron elementos virtuales de estudio OVA o ámbitos virtuales de Estudio EVA Análisis de Desempeño o Evaluación de estudiantes: Trabajos en los cuales la utilización de técnicas de minería de datos se orientó a investigar el desempeño de los alumnos o las evaluaciones completados a estos a lo largo de los tutoriales presenciales o virtuales.
- **Generación de Recomendaciones:** dominio en el cual se asociaron los trabajos que generaron sugerencias a ser utilizadas a lo largo del desarrollo educativo.
- **Identificación de patrones de aprendizaje:** análisis de datos de ámbitos académicos que permitieron detectar patrones de estudio a ser usados por las instituciones educativas o instructores.
- **Identificación de patrones en estudiantes:** análisis de datos de ámbitos académicos que permitieron detectar patrones en los alumnos.
- **Predicción relacionada a estudiantes:** en este dominio están los trabajos que permitieron hacer conjeturas similares a los alumnos, conjeturas en notas finales, desarrollo, accionar en tutoriales concreto, etc.

En la siguiente tabla se presentan los trabajos relacionados a cada uno de los dominios presentados anteriormente.

Tabla 3
Clasificación de trabajos por dominios presentes en la educación.

Dominio	No. de Publicaciones	Autores Relevantes	Años con publicaciones
Análisis de Deserción o Retención	6	Marquez-Vera C., Romero C., Ventura, S., Bayer J., Kovačić, Z. J	Entre el 2010 y 2015
Análisis de OVA o EVA	26	Romero C., Ventura S., De Bra P., Karampiperis P., Sampson D.	Entre el 2001 y 2015
Análisis de Rendimiento o Evaluación de estudiantes	17	Bhardwaj B., Pal S., Romero C., Espejo, P., Zafra A., Romero R., Ventura S.	Entre el 2007 y 2015
Generación de Recomendaciones	14	Wang F.H., Shao H.M., Vialardi C., Brav J. Shafti L., Ortigosa A.	Entre el 2004 y 2014
Identificación de patrones de aprendizaje	25	Zaiane O., Garcia E., Romero C., Ventura S., Castro C., Ha S., Bae S., Park S.	Entre el 2000 y 2015
Identificación de patrones en estudiantes	23	Baradwaj B., Pal S., Mor E., Minguillón J.	Entre el 1995 y 2015
Predicción relacionada a estudiantes	20	Oladokun V.O., Adebajo A.T., Charles-owaba O.E,	Entre el 1993 y 2015

2.1.2.9. Gestión y dirección de proyectos

La gestión de proyectos es un área de conocimiento que tiene mucho tiempo madurando, hasta convertirse hoy en un estándar y mejores prácticas bajo varias compilaciones. Según Valle et al (2010), citados por (Mustaro & Rossi, 2013) la construcción de las pirámides de Egipto, la gran Muralla China y el Coliseo en Roma representan ejemplos de los esfuerzos del hombre que deben haber sido regidos por principios de gestión de proyectos. Asimismo, Henry Gantt, famoso por su diagrama de Gantt, fue el fundador de la planificación y el control, mientras que Henry Fayol fue el fundador de la predicción y la planificación, la organización, la delegación, la coordinación y el control (Stevens, 2002; Morgen, 2003; Morris et al 2006 citados por Ljevo & Vukomanović, 2014). Tanto Gantt y Fayol eran estudiantes de las teorías de la administración científica de Frederick Winslow Taylor, cuyo trabajo fue el precursor de herramientas modernas de gestión de proyectos, incluyendo la estructura de desglose del trabajo (WBS) y la asignación de recursos (Ljevo & Vukomanović, 2014). Estos autores contribuyeron en gran medida a las prácticas de gestión de proyectos moderna, y que a comienzos del siglo XX fue donde se establecieron las teorías, metodologías, marcos destinados a la gestión de proyectos (Mustaro & Rossi, 2013)

Cada marco de trabajo, guía, framework, buenas prácticas o método para la gestión de proyectos presentado tiene sus propias características y ventajas, la cual presentan casos de éxito en la gestión de proyectos, por tal razón son considerados en su entorno como tal. Sin embargo, existe una clasificación o agrupación de estos marcos de trabajo basados en la importancia de implantación como: estándares internacionales, guías o normas y métodos (Palacios, 2014), tal como se aprecia en la siguiente Figura N° 14.

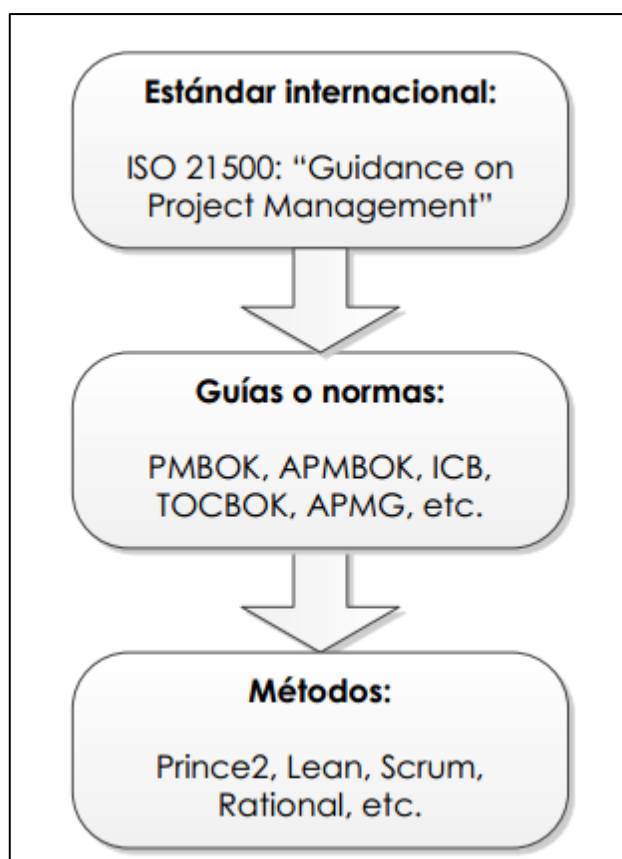


Figura 11. Diagrama de organización para modelos adaptado de Figuerola (2013) citado por Palacios (2014)

El interés por los proyectos y su utilización para la implementación de la estrategia organizacional ha aumentado vertiginosamente en los últimos años a nivel mundial, por tal motivo el contar con un sistema de gestión de proyectos que responda a las exigencias de adaptabilidad y flexibilidad se hace necesario, así como la eficiencia en la asignación de recursos, para tener una gestión éxitos (Solarte-Pazos & Sánchez-Arias, 2014). Asimismo, se ha considerado que el éxito en los proyectos radica en el cumplimiento de coste, tiempo y requisitos (Igartua Lopez J. I, Errasti Lozares N, & Zabaleta Etxebarria, 2012).

Los autores Zabaleta, Igartua, & Errasti (2012), señalan los principales estándares en gestión de proyectos a PMBOK, APMBOK, BS 6079 (British Standard Guide to Project

Management), IPMA, ISO 10006 y P2M (Project & Program Management for Enterprise Innovation), de acuerdo a su estudio realizado por focalizarse netamente en el proyecto.

Sin embargo, para la implantación de alguna de estas buenas prácticas para la gestión de proyectos, depende de los conocimientos y habilidades del Director del Proyecto, el cual debe tener conocimiento de estándares y mejores prácticas en la gestión de proyectos.

2.1.2.10. PMBOK

La Guía de Fundamentos para la Dirección de Proyectos (Guía del PMBOK), contiene un conjunto de estándares (buenas prácticas) reconocidos a nivel mundial. Además, promueve el uso de un vocabulario común para el uso y aplicación de los conceptos de la dirección de proyectos dentro de la profesión de la dirección de proyectos (Project Management Institute, Inc, 2013). Así la definición de proyecto establecida ya como un estándar es el siguiente: “Un proyecto es un esfuerzo temporal que se lleva a cabo para crear un producto, servicio o resultado único. La naturaleza temporal de los proyectos implica que un proyecto tiene un principio y un final”.

El PMI fue fundado en 1969 en Estados Unidos de Norteamérica, con el objetivo de servir a los intereses de la industria de la gestión de proyectos (Matos & Lopes, 2013), y se ha convertido en una de las principales organizaciones profesionales en cuanto a dirección de proyectos se refiere, desde su fundación el PMI ha sido logrando posicionarse como un organismo de normalización, tal como se aprecia en la siguiente tabla, su proceso de crecimiento hasta la fecha.

Tabla 4
Origen e historia del PMBOK

Año	Evento
1969	Fundación del PMI
1981	El PMI autorizó el desarrollo de lo que fue el comienzo de la Guía del PMBOK.
1983	Reporte especial en Ética, Normas y Acreditación del PMI. Fue parte de los estándares del cuerpo de conocimientos de la gestión de proyectos.
1987	El estándar PMBOK fue publicado.
1996	Publicación de la Primera Edición de La Guía de los Fundamentos para la Dirección de Proyectos (Guía PMBOK).
1999	PMI es acreditado como una Organización de Normalización por el Instituto Nacional Estadounidense de Estándares (ANSI, por sus siglas en inglés: American National Standards Institute).
2000	La Guía PMBOK – Edición 2000 (Primera Edición) fue publicado y es reconocido como estándar ANSI/PMI 99-001-2000
2004	La Guía PMBOK - Edición 2004 (Tercera Edición) fue publicado y es reconocido como estándar ANSI/PMI 99-001-2004
2008	La Guía PMBOK - Edición 2008 (Cuarta Edición) fue publicado y es reconocido como estándar ANSI/PMI 99-001-2008
2013	La Guía PMBOK - Edición 2013 (Quinta Edición) fue publicado y es reconocido como estándar ANSI/PMI 99-001-2013

Fuente: Adaptado de Assaff (2007) y Matos & Lopes (2013)

Sin embargo, para que el PMI, pueda establecer los estándares para la dirección de proyectos, tuvo que madurar en el tiempo, en base a las buenas prácticas de todos lo que expertos, especialistas y profesionales dedicados a la dirección de proyectos, los mismos que han contribuido para el desarrollo del PMBOK. El cual es aplicado activamente en la gestión de proyectos de la industria del software y la industria de la construcción (Project Management Institute, 2008 citado por Matos & Lopes, 2013).

a) Ciclo de vida del proyecto

El ciclo de vida de un proyecto es referido a las fases por la que atraviesa un proyecto, desde que inicia hasta su término. Estas fases son generalmente secuenciales y sus nombres y números se determinan en función de las necesidades de gestión y control de la organización, la naturaleza propia del proyecto y su área de aplicación. Así como también acotadas en el tiempo, con un inicio y un final o punto de control; sin embargo, mientras que el proyecto recorre su ciclo de vida, los entregables

específicos y las actividades que se llevan a cabo variarán ampliamente dependiendo del proyecto. Esto indica que los proyectos varían en tamaño y complejidad, aun así, pueden configurarse dentro de una estructura de inicio del proyecto, organización y preparación, ejecución del trabajo y cierre del proyecto (Project Management Institute, 2013).

Tal como se ilustra en la siguiente Figura N° 12.

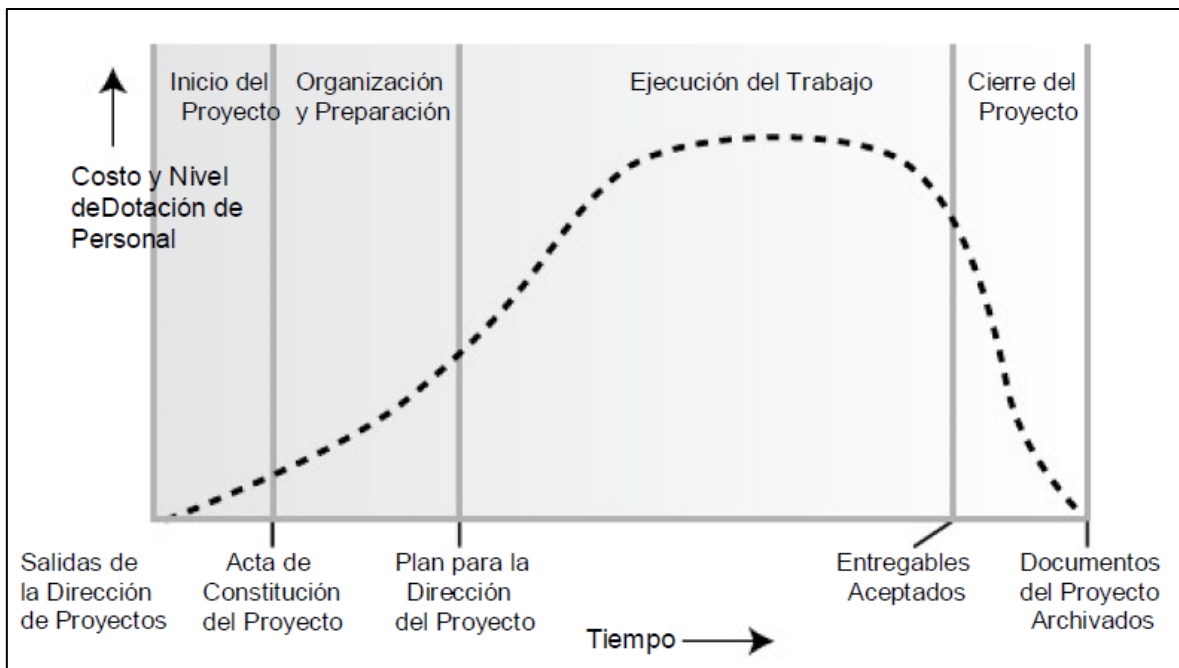


Figura 12. Ciclo de vida del proyecto según costos y personal adaptado de Project Management Institute (2013)

b) Procesos de la dirección de proyectos

Se acuerdo al Project Management Institute (2013), declara que la dirección de proyectos es la aplicación de conocimientos, habilidades, herramientas y técnicas a las actividades del proyecto para cumplir con los requisitos del mismo; y requiere de una gestión eficaz de los procesos de dirección de proyectos. Esto se logra mediante la aplicación e integración adecuada de los 47 procesos de la dirección de proyectos, agrupados de manera lógica, categorizados en 5 grupos de procesos (ver Figura N° 16).

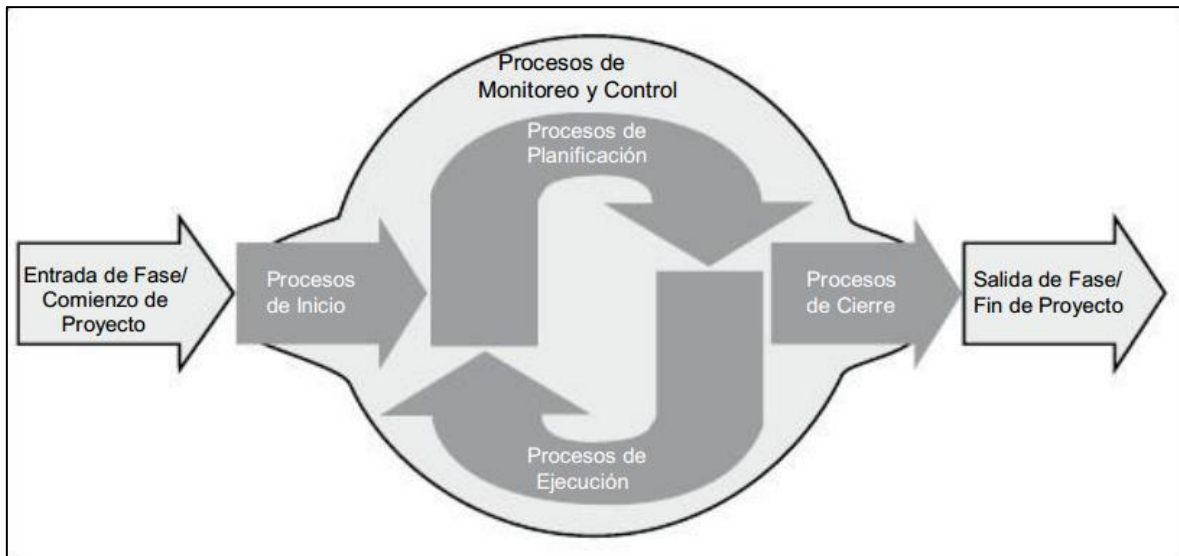


Figura 13. Grupo de Procesos de la Dirección de Proyectos adaptado de Project Management Institute (2013)

Estos procesos de la dirección de proyectos son los procesos de **inicio**, este grupo de procesos son aquellos realizados para definir un nuevo proyecto o nueva fase de un proyecto al obtener una autorización para iniciar el proyecto o fase. Luego siguen los procesos de **planificación**, este grupo de procesos son aquellos requeridos para establecer el alcance del proyecto, refinar los objetivos y definir el curso de acción requerido para alcanzar los objetivos propuestos del proyecto. Continúa los procesos de **ejecución**, este grupo de procesos son aquellos realizados para completar el trabajo definido en el plan para la dirección del proyecto a fin de satisfacer las especificaciones del mismo. También están los procesos de **monitoreo y control**, este grupo de procesos son aquellos requeridos para rastrear, revisar y regular el progreso y el desempeño del proyecto, para identificar áreas en las que el plan requiera cambios y para iniciar los cambios correspondientes, y finalmente los procesos de **cierre**, finalmente este grupo procesos son aquellos realizados para finalizar todas las actividades a través de todos los grupos de procesos, a fin de cerrar formalmente el proyecto o una fase del mismo. (Project Management Institute, Inc, 2013)

c) Áreas de conocimiento

La versión del PMBOK más reciente considera 10 áreas de conocimiento, un área de conocimiento más que la versión anterior que solo consideraba 9, el área que se ha agregado fue la Gestión de Interesados (Project Management Institute, 2013). Estas áreas son las siguientes:

Gestión de Integración del Proyecto

Esta área incluye los procesos (6) y actividades necesarias para identificar, definir, combinar, unificar y coordinar los diversos procesos y actividades de dirección del proyecto dentro de los Grupos de Procesos de la Dirección de Proyectos.

Gestión del Alcance del Proyecto

Esta área incluye los procesos (6) necesarios para garantizar que el proyecto incluya todo el trabajo requerido y únicamente el trabajo para completar el proyecto con éxito. Básicamente se define y controla que se incluye y que no se incluye en el proyecto.

Gestión del Tiempo del Proyecto

Esta área incluye los procesos (7) requeridos para gestionar la terminación en plazo del proyecto.

Gestión de los Costos del Proyecto

Esta área incluye los procesos (4) relacionados con planificar, estimar, financiar, obtener financiamiento, gestionar y controlar los costos de modo que se complete el proyecto dentro del presupuesto aprobado.

Gestión de la Calidad del Proyecto

Esta área incluye los procesos (3) y actividades de la organización ejecutora que establecen las políticas de calidad, los objetivos y las responsabilidades de calidad para que el proyecto satisfaga las necesidades para las que fue acometido.

Gestión de los Recursos Humanos del Proyecto

Esta área incluye los procesos (4) que organizan, gestionan y conducen al equipo del proyecto.

Gestión de las Comunicaciones del Proyecto

Esta área incluye los procesos (3) requeridos para asegurar que la planificación, recopilación, creación, distribución, almacenamiento, recuperación, gestión, control, monitoreo y disposición final de la información del proyecto sean oportunos y adecuados.

Gestión de los Riesgos del Proyecto

Esta área incluye los procesos (6) para llevar a cabo la planificación de la gestión de riesgos, así como la identificación, análisis, planificación de respuesta y control de los riesgos del proyecto.

Gestión de las Adquisiciones del Proyecto

Esta área incluye los procesos (4) necesarios para comprar o adquirir productos, servicios o resultados que es preciso obtener fuera del equipo del proyecto.

Gestión de los Interesados del Proyecto

Esta área incluye los procesos (4) necesarios para identificar a las personas, grupos u organizaciones que pueden afectar o ser afectados por el proyecto, para analizar las expectativas de los interesados y su impacto en el proyecto, y para desarrollar estrategias de gestión adecuadas a fin de lograr una participación eficaz de los interesados en las decisiones y en la ejecución del proyecto.

CAPITULO III:

MÉTODO

3.1. Tipo de Investigación

Básica

Ya que presenta como finalidad la creación de conocimiento, es decir, la obtención y recopilación de información para la creación de un nuevo modelo para el desarrollo de proyectos de minería de datos.

Aplicada

La investigación aplica el nuevo modelo a un caso del mundo real.

Diseño de Investigación

El diseño de esta investigación es cuasi experimental, el cual trabaja con grupos ya formados, no aleatorizados, por tanto, su validez interna es pequeña porque no hay control con las variables extrañas. Estos diseños se aplican a situaciones reales en los que no se pueden formar grupos aleatoriamente, pero pueden manipular la variable experimental. (Hernández et. al. citado por Ñaupás, et. al. 2014)

3.2. Población y muestra

Población

Expertos en aplicación de minería de datos en organizaciones académicas.

Muestra

Se ha considerado a 30 especialistas o expertos en minería de datos de diferentes organizaciones. Considerando que la población es menor a 50 individuos, la muestra es igual a la población (Castro, 2003). Por lo que la cantidad mínima representativa para validar el estudio será de 30 expertos.

3.3. Operacionalización de las variables.

Tabla 5
Operacionalización de variables

VARIABLE	DIMENSIONES	INDICADORES
Modelo	Existencia	Si-No
Proceso de Aplicación de Minería de Datos	Fases del proceso	Porcentaje de Cumplimiento Metodológico
	Gestión del proyecto de minería de datos	Porcentaje de Cumplimiento de la Gestión de Proyecto
	Actividades específicas de la minería de Datos	Porcentaje de cumplimiento en actividades de Desarrollo de MD

Nota: Fuente Elaboración propia

3.4. Instrumentos

Encuesta – Cuestionario

3.5. Procedimiento

Aplicación de encuesta

Tabulado

Análisis de Datos

3.6. Análisis de Datos

Diferencia de Medias

Análisis descriptivo

**CAPITULO IV:
DESARROLLO DEL
NUEVO MODELO**

4.1. Elaboración del Modelo

Como se menciona anteriormente la metodología CRISP-DM no hace distinción entre procesos, sino que posee un único proceso en el cual se definen todas las tareas dividiéndolas en fases, ya sea las relacionadas con la administración como las relacionadas con el desarrollo.

Esta división es necesaria ya que ambos procesos deben ser ejecutados en forma concurrente, por un lado, se debe desarrollar el proyecto propiamente dicho y al mismo tiempo se deben ejecutar tareas de administración del proyecto para controlar su avance, realizar correcciones y recolectar datos para futuros proyectos.

La reestructuración de estos dos procesos en el desarrollo de proyectos de Explotación de Información brindará mayor claridad a las tareas de administración ya que, como se menciona anteriormente, las tareas de administración que se encuentran mencionadas en las metodologías evaluadas se encuentran dentro de un mismo proceso de desarrollo.

Esta división no se limita a separar las tareas existentes en las metodologías, sino que se construye el nuevo proceso basándose en el proceso de administración de proyectos, lo cual enriquece el proceso de administración de proyectos. A este proceso se lo adecuó para responder a las exigencias de los proyectos de Explotación de Información.

La incorporación del proceso de administración del proyecto, y su concepción como un proceso distinguible del de desarrollo, constituye una aportación a la gestión de los proyectos de Explotación de Información, ya que ninguna de las metodologías evaluadas, incluyendo a CRISP-DM, hace una separación clara entre las tareas que se deben llevar a cabo para la administración del proyecto y las que se llevan a cabo para el desarrollo del mismo.

Se presenta MoPro-DM, un modelo de procesos basado en PMI y CRISP-DM, que integra un conjunto de roles, documentos, y actividades, para desarrollar proyectos de minera de datos.

4.1.1. Representación Conceptual del nuevo modelo

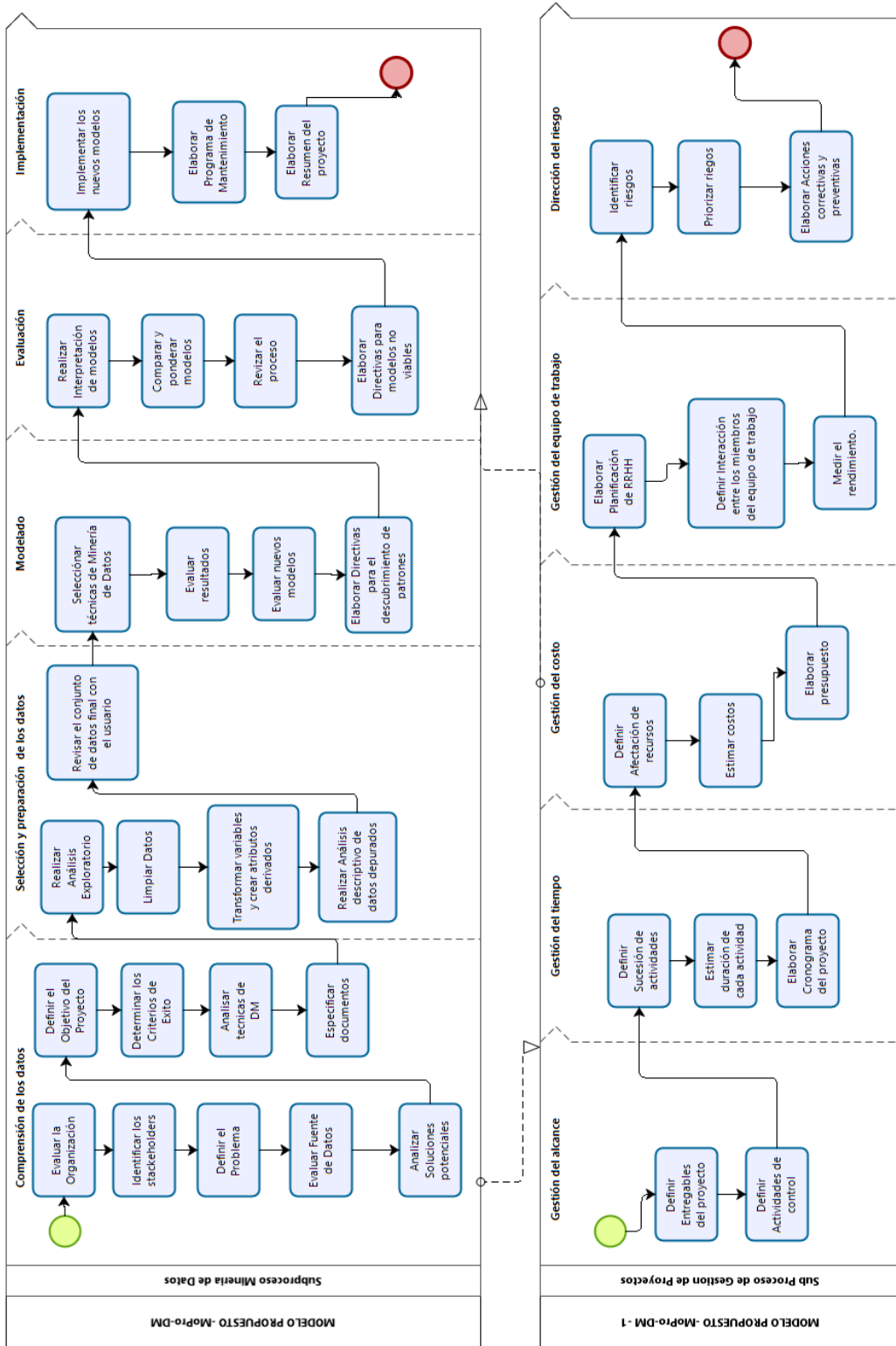


Figura 14. Proceso del propuesto

4.1.2. Roles

Después de consultar varias literaturas generales de desarrollo de sistemas que hacen indicaciones de que los proyectos de minería de datos requieren más personal calificado y especializado que cubran los diversos aspectos y requisitos de habilidades que surgen durante la ejecución de un proyecto de minería de datos se hizo más obvio. Los siguientes recursos humanos han sido identificados como necesario para abordar e implementar un proyecto de minería de datos con éxito.

Todos los recursos humanos identificados han sido tomados de (Ganesh, Han, Kumar, Shekhar, & Srivastava, 1996), (Lee & Kerschberg, 1998), (Hofmann & Tierney, 2009) :

Tabla 6
Roles de Gestión de Proyecto

Rol	Competencias
EDM-Experto del Dominio	Un experto en el dominio es un experto en la materia con antecedentes relevantes, experiencia o experiencia en temas específicos.
EDM-Gestor del Proyecto	Especialista técnico encargado de coordinar el proyecto de Minería de Datos con los demás participantes
EDM- Arquitecto de Datos	Especialista técnico experto en la implementación del sistema de Data Warehouse (DWH), por lo que debe conocer el dominio de los datos disponibles en los Data Marts que lo conforman. En caso de que existan orígenes de datos adicionales al DWH relevantes para el análisis, será la persona que tenga el criterio para decidir si estos orígenes son convenientes de ser incorporados al DWH y las políticas que definirán la implementación del nuevo Data Mart.
EDM- Administrador de Base de Datos	especialista técnico, con conocimiento de las bases de datos y estructuras disponibles en los sistemas de la universidad, este rol podrá ser desempeñado por más de una persona dado el número de sistemas manejado por la universidad.
EDM- Experto en Minería de Datos	Es el responsable de la ejecución de las actividades de análisis de datos
EDM- Analista de Negocios	Especialista experto en la temática de negocio sobre la cual el problema o requerimiento se desarrolla. Igualmente, dada la extensión del problema planteado podría ser una o más personas.

4.1.3. Modelo propuesto de aplicación de Minería de Datos para el sector educativo.

Se han seleccionado 3 aspectos basados en cada fase del proceso de minería de datos (Ver Figura 14) para este análisis: Modelo, Sub proceso de minería de datos y sub proceso de la dirección de proyectos.

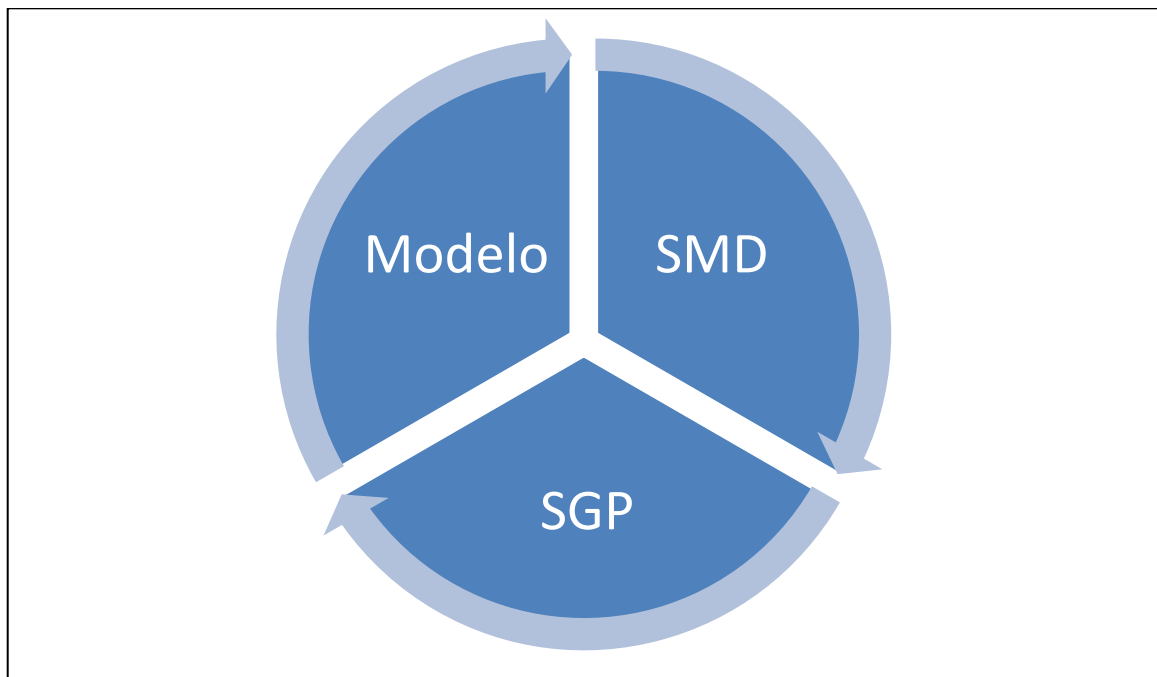


Figura 15. Modelo propuesto.

4.1.4. Nivel de detalle de las actividades

Se debe evaluar el grado de profundidad con el que se definen las actividades específicas que forman parte de cada etapa del proceso e incluir una guía de cómo ejecutarlas.

La importancia de este aspecto radica en el nivel de abstracción (fases) y la capacidad de definir un conjunto de actividades específicas que secuencien el trabajo en cada una de ellas. En la Tabla 7, evidenciaremos las características a evaluar es este aspecto.

Tabla 7
Características para el nivel de detalle de las actividades

CARACTERÍSTICA	DESCRIPCIÓN	RESULTADO ESPERADO
Definición de actividades específicas	Se debe diseñar un conjunto de actividades específicas que especifican el trabajo a realizar.	Definición de actividades específicas de acuerdo al nivel de abstracción.
Especificación de pasos a seguir para cada actividad	Con el fin de mitigar conjeturas erróneas, se debe resumir los pasos necesarios para ejecutar cada una de las actividades.	Detalle de las actividades específicas.
Definición de las entradas de cada actividad (Input)	Permite al equipo de trabajo conocer las condiciones de inicio y los requisitos correspondientes	Prerrequisitos necesarios para cada actividad
Definición de salidas de cada actividad (Output)	Define los entregables de cada actividad es decir, representan los resultados adquiridos posteriores a la ejecución.	Entregables (resultados) de cada actividad.

Entrega de una guía de buenas prácticas para cada una de las actividades.

Es imprescindible entregar al usuario (sin importar el nivel de conocimiento) sugerencias que aminoren los inconvenientes que pudieran trascender durante la ejecución.

Enumeración de consejos para la ejecución de cada actividad.

4.1.5. Actividades específicas de Minería de datos

En este aspecto se evalúan las actividades principales que imperativamente deben estar presentes en cada fase del proceso de minería de datos, de la misma manera se analiza la incorporación de actividades adicionales y relevantes para la organización. De acuerdo a lo anterior expuesto, evaluaremos sus características en la siguiente tabla.

Tabla 8

Evaluación de características con respecto a las fases del proceso de minería de datos

FASE	CARACTERÍSTIC	DESCRIPCIÓN	RESULTADO
	A	N	ESPERADO
COMPRENSIÓN DE LOS DATOS	Evaluación de la organización	Se evidencian objetivos del negocio, estructura de la organización y resumen de actividades que se desarrollan en la organización.	Evaluación general de la organización.
	Identificación de stakeholders	Se identifica a todas aquellas personas implicadas en el proyecto que tienen algún de interés o intervienen en el mismo.	Detalle de los involucrados en el proyecto.
	Definición del problema	Permite evidenciar el tipo patrones para la	Definición de problemática organizacional

	búsqueda y determina los objetivos del proyecto y las actividades a ejecutar	
Evaluación de fuentes de datos	Estudio de las fuentes de datos como entrada para la explotación de información, análisis de estructura y accesibilidad para planificación y construcción del conjunto de datos.	Evaluación de fuentes de datos.
Análisis de soluciones potenciales	Análisis previo de las posibles soluciones al problema con su justificación correspondiente.	Análisis de las alternativas de solución.
Definición de objetivos del proyecto	Detalle de los objetivos del proyecto tanto desde el enfoque organizacional	Determinación de objetivos.

	como del técnico.	
Determinación de criterio de éxito	Especificación de cuándo el proyecto ha alcanzado los resultados adecuados que satisfagan los objetivos planteados.	Definición de criterios de éxito.
Análisis de técnicas de DM	Reside en el análisis de técnicas procedentes que permitan alcanzar los objetivos del proyecto.	Evaluación de técnicas de minería datos.
Especificación de documentos (entregables)	Definición de documentación con la que los usuarios manejaran los resultados del proyecto, permite comprender cómo los nuevos modelos se integran en los procesos de negocio.	Especificación, difusión y uso de nuevos patrones.

SELECCIÓN Y PREPARACIÓN DE LOS DATOS	Análisis Exploratorio	Al seleccionar y	Análisis exploratorio/descriptivo de los datos.
		imprescindible familiarizarse con los mismos; se debe estudiar la distribución y comportamiento de las variables, identificando también tareas de limpieza.	
	Limpieza de datos	Los datos recolectados inicialmente, suele contener errores e incluso estar ausentes., la limpieza de datos permite no caer en el fenómeno GIGO (Garbage In, Garbage Out).	Detalle de actividades de limpieza de datos
	Transformación de variables y creación de atributos derivados	Los datos obtenidos inicialmente, suelen ingresar en diferente formato, lo que no permite	Detalle de actividades para la transformación de los datos.

realizar el
 proceso de
 minería, es
 necesario
 realizar la
 transformación
 de las variables
 y el cálculo de
 atributos
 derivados (su
 valor surge de
 otros campos)

Análisis descriptivo
 de datos depurados

Con los datos en
 formatos
 correctos y
 depurados, el
 análisis
 descriptivo del
 conjunto de
 datos permite la
 modelización e
 interpretación.

Análisis descriptivo
 final del conjunto de
 datos final.

Revisión del
 conjunto de datos
 final con el usuario

El objetivo de
 todo el proceso,
 es que los datos
 reflejen de la
 mejor manera la
 realidad, es
 necesario
 corroborar que
 se vislumbre
 todos los casos
 posibles para

Detalle de la revisión
 del conjunto de datos
 con el usuario.

evitar omitir
valores de las
variables.

MODELADO	Selección de técnicas de Minería de Datos	No todas las técnicas son aplicables en todos los casos, se debe evaluar los factores de influencia para la selección, por ejemplo se debe tomar en cuenta el tamaño del conjunto de datos o la naturaleza de las variables.	Análisis de técnicas de minería a implementar para la creación de los modelos.
	Evaluación de resultados	Se debe definir cómo se dividirá el conjunto de datos para las pruebas de los modelos, adicionalmente se debe establecer criterios para la ponderación de los resultados.	Definición de criterios de evaluación.

	Evaluación de nuevos modelos	Se debe evaluar los nuevos modelos obtenidos en función de los criterios de evaluación. Esta evaluación debe contener también la descripción de los modelos.	Evaluación detallada de nuevos modelos de datos.
	Directivas para el descubrimiento de patrones	Debido a que en todos los casos se obtiene resultados positivos, es primordial verificar las actividades alternativas que propone la metodología seleccionada.	Definición de alternativas para lograr descubrir patrones en el conjunto de datos.
EVALUACIÓN	Interpretación de modelos	Se debe analizar los modelos obtenidos y su adecuación a los objetivos organizacionales, este análisis debe interpretarse en	Detalle de la interpretación de modelos en función de los objetivos del negocio.

	función de la situación del negocio, determinando si los mismos resultan son útiles desde una perspectiva organizacional	
Comparación y ponderación de modelos	Posterior al análisis de los modelos, se debe ponderar de manera individual para determinar cuáles son los más sólidos y que mejor se adaptan a los objetivos planteados.	Determinación de modelos adaptables a los objetivos.
Revisión del proceso	Se debe realizar una revisión de todo el proceso, analizando la correspondencia entre los objetivos y los resultados obtenidos, asegurando así que no se omitió	Revisión del proceso de manera macro

ningún tema importante.

Directivas para modelos no viables Existen casos en los que a pesar de llevar todo el proceso de manera correcta, los modelos obtenidos no son viables porque no aportan conocimiento nuevo, no son factibles o su interpretación es deficiente, la metodología debe proponer actividades para retomar el proceso.

IMPLEMENTACIÓN N Implementación de nuevos modelos Desarrollo de un plan para especificar las actividades para la difusión y generalización del uso de la información obtenida Planificación de programa de difusión organizacional.

Programa de Mantenimiento	<p>Después de la implementación de los modelos, los modelos deben ser comprobados de manera periódica para verificar su validez; el entorno es variable en el tiempo de acuerdo sus condiciones.</p>	Desarrollo de plan de mantenimiento.
Resumen del proyecto	<p>La metodología debe entregar un resumen del proyecto donde se evidencie la evolución del proceso y los resultados obtenidos. También se debe plasmar buenas y malas prácticas efectuadas durante el transcurso del proyecto</p>	Reporte con resumen del proyecto.

4.1.6. Actividades de dirección del proyecto

Si bien estas actividades están enfocadas en la administración de otros aspectos como: tiempo (cronograma), costo (presupuesto) o el riesgo; también tienen por objetivo incrementar la probabilidad de que el proyecto culmine en el tiempo estimado, con el presupuesto asignado y de manera exitosa.

De manera general estas actividades se dividen en dos grupos que permiten mantener de manera macro una gestión adecuada:

- **Planificación:** circunscriben la identificación de tareas, estimación de tiempo de ejecución y de recursos, y definición de trayectoria del proyecto.

- **Control:** contiene actividades de seguimiento del estado del proyecto en concordancia con la planificación inicial.

Independientemente del tipo de proyecto y la industria, existen estándares que instauran las áreas a administrar para una gestión conveniente de proyectos; PMBOK (Project Management Institute, 2016), define cinco áreas esenciales para la dirección: gestión del alcance, gestión del tiempo, gestión del costo, gestión del equipo de trabajo y gestión del riesgo.

4.1.6.1. Gestión del alcance

Consiste en la planificación y control de todo lo que se ejecutará en el transcurso del proyecto, define los entregables que serán incluidos y el alcance del trabajo a realizar. En la siguiente tabla se resume las características que debe tener cada área de acuerdo a las características de esta fase.

Tabla 9
Características de la Gestión del Alcance

CARACTERÍSTICA	DESCRIPCIÓN	RESULTADO ESPERADO
Entregables del proyecto	De acuerdo a la metodología optada, el equipo de trabajo debe seleccionar y definir cuáles son los entregables que forman parte del proyecto, esto permite organizar de forma escalonada el trabajo a ser ejecutado.	Detalle de entregables
Actividades de control	Permite dar seguimiento al final de cada fase a los cambios en la planificación, tanto de actividades como entregables.	Revisión del avance de cada fase

4.1.6.2. Gestión del tiempo

Las actividades de esta área tienen como objetivo finalizar el proyecto de manera exitosa en el tiempo estipulado en la planificación inicial. A continuación, en la Tabla 13 se detalla las características.

Tabla 10
Características de la Gestión del Tiempo

CARACTERÍSTICA	DESCRIPCIÓN	RESULTADO ESPERADO
Sucesión de actividades	Posterior a la definición de las actividades, se analiza la relación y secuencia de las mismas para la ejecución en el proyecto.	Definición de secuencia de actividades por fase.
Estimación de duración de cada actividad	Permite evaluar la duración de cada actividad en función de los recursos humanos disponibles y su skill correspondiente.	Evaluación de tiempo de ejecución por actividad.
Cronograma del proyecto	Se debe realizar un cronograma donde se define fechas de inicio y fin para actividades principales dentro de cada fase.	Cronograma completo del proyecto

4.1.6.3. Gestión del costo

Uno de los aspectos más importantes es el control de los costos, esta área debe realizar la planificación y control mediante estimación de costos y la generación de presupuesto del proyecto. Como se evidencia en la Tabla 14, esta gestión también toma en cuenta de la afectación de los recursos.

Tabla 11
Características de la Gestión de Costos

CARACTERÍSTICA	DESCRIPCIÓN	RESULTADO ESPERADO
Afectación de recursos	Permite aprobar los recursos (materiales y humanos) necesarios para la ejecución de las actividades.	Estimación de recursos para cada actividad.
Estimación de costos	Se debe evaluar los diversos escenarios y el tiempo sobre los que se ejecuta el proyecto.	Estimación de costos para cada actividad.
Elaboración de presupuesto	Se debe analizar todos los costos estimados por actividad para obtener una línea base del total y medir la aptitud del proyecto.	Presupuesto general del proyecto

4.1.6.4. Gestión del equipo de trabajo

Esta área es responsable de los procesos de recursos humanos, se enfocan en el equipo de trabajo que han sido asignado roles y responsabilidades en las actividades del proyecto. Se detalla sus características en la siguiente tabla.

Tabla 12

Característica de Gestión del equipo de trabajo

CARACTERÍSTICA	DESCRIPCIÓN	RESULTADO ESPERADO
Planificación de RRHH	Identificación y asignación de roles y responsabilidades del equipo de trabajo.	Determinación de roles y responsabilidades
Interacción entre los miembros del equipo de trabajo	Mantener una comunicación productiva permite que las reuniones periódicas fluyan y que la resolución grupal de conflictos sea más eficiente.	Reuniones periódicas de seguimiento entre los miembros del equipo de trabajo.
Medición de rendimiento.	Permite dar seguimiento de manera indirecta al avance del proyecto y evidencia la aplicación políticas de comunicación y motivación	Métricas de rendimiento de equipo de trabajo

4.1.6.5. Dirección del riesgo

Se considera riesgo, a un evento o condición fortuita que al presentarse tiene un efecto (positivo o negativo) sobre uno o más objetivos del proyecto, su impacto se mide en función de las consecuencias. Esa área tiene como objetivos: identificar y priorizar los principales riesgos, evaluar el impacto y dar el seguimiento correspondiente, tal cual se evidencia en la Tabla siguiente.

Tabla 13
Característica de la Dirección del Riesgo

CARACTERÍSTICA	DESCRIPCIÓN	RESULTADO ESPERADO
Identificación de riesgos	Evidenciar y detallar posibles riesgos que podrían afectar al proyecto.	Detalle de riesgos
Priorización de riesgos	Cada riesgo debe tener su probabilidad de ocurrencia e impacto en el proyecto tomando en cuenta la interacción entre los mismos.	Detalle de priorización de riesgos
Acciones correctivas y preventivas	Determinar estrategias y responsables de prevención del riesgo y de respuesta.	Plan de acción ante riesgos.

CAPITULO V: RESULTADOS

4.1. resultados Específicos

Resultados de Pre-Prueba y Post para los KPI₁, KPI₂, KPI₃, son mostrados en la tabla 14:

Tabla 14
Resultados por cada indicador (KPI)

	KPI1: Porcentaje de Cumplimiento Metodológico		KPI2: Porcentaje de Cumplimiento de la Gestión de Proyecto		KPI3: Porcentaje de cumplimiento en actividades de Desarrollo de MD	
	KP1_PRE	KP1_POST	KP2_PRE	KP2_POST	KP3_PRE	KP3_POST
1	79	90	87	93	85	93
2	83	97	88	92	84	91
3	84	96	84	94	82	93
4	88	93	87	93	89	95
5	86	96	88	92	85	92
6	83	93	86	94	90	92
7	84	97	84	90	87	95
8	84	99	80	94	87	90
9	90	96	84	92	86	92
10	87	91	83	93	86	95
11	89	97	84	90	88	91
12	85	98	82	91	86	89
13	85	97	86	94	88	94
14	86	95	79	92	89	93
15	89	95	81	93	82	93
16	83	92	82	92	84	93
17	85	98	83	92	83	92
18	88	94	81	92	85	91
19	87	92	82	93	82	92
20	82	91	87	91	84	93
21	83	98	80	94	90	93
22	86	95	87	95	87	93
23	86	98	85	94	88	93
24	88	91	90	90	89	92
25	87	96	86	90	84	92
26	82	95	82	93	84	90
27	86	94	85	93	84	92
28	87	94	85	90	86	94
29	84	97	86	91	81	89
30	85	99	81	90	88	91

4.2. Análisis de Resultados Descriptivos

En las siguientes tablas, se muestra los resultados de la estadística descriptiva de la Pre-prueba y Post-prueba. Además, se resalta los valores de los KPIs medidos, en la Post Prueba, que son mejores (mayores o menores) que los KPI promedio en la Post Prueba. A continuación, se realiza un análisis detallado de los datos de cada una de las tablas

A. KPI1: Porcentaje de Cumplimiento Metodológico

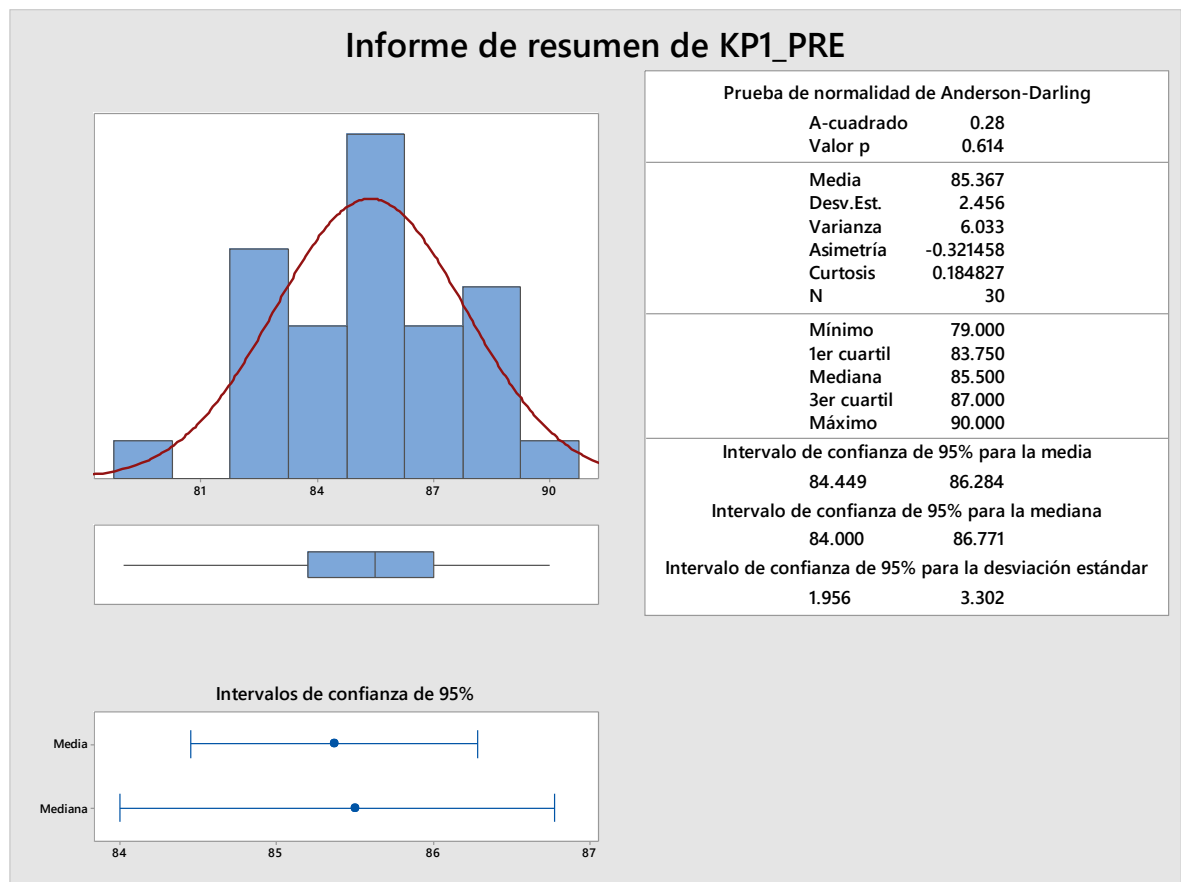
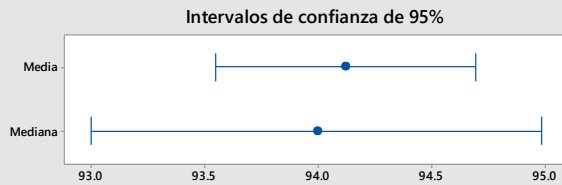
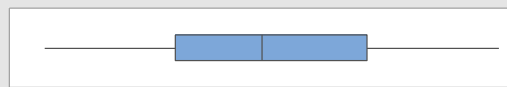
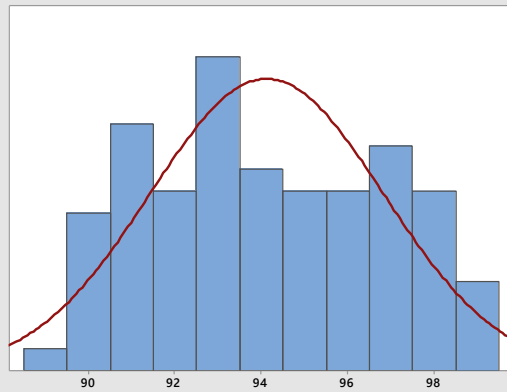


Figura 16. Estadística Descriptiva KPI1_PrePrueba

De la figura anterior se puede ver que los datos tienen una distribución normal puesto que el valor de $p > 0.05$ con una media de 85% y una desviación estándar de 2.4.

Informe de resumen de KP1_POST



Prueba de normalidad de Anderson-Darling	
A-cuadrado	1.02
Valor p	0.011
Media	94.122
Desv. Est.	2.698
Varianza	7.282
Asimetría	0.12959
Curtosis	-1.04714
N	88
Mínimo	89.000
1er cuartil	92.000
Mediana	94.000
3er cuartil	96.452
Máximo	99.482
Intervalo de confianza de 95% para la media	
	93.551 94.694
Intervalo de confianza de 95% para la mediana	
	93.000 94.986
Intervalo de confianza de 95% para la desviación estándar	
	2.350 3.169

Figura 17. Estadística Descriptiva KPI1_PostPrueba

De la figura anterior se puede ver que los datos tienen una distribución normal puesto que el valor de $p > 0.05$ con una media de 85% y una desviación estándar de 2.4.

B. KPI2: Porcentaje de Cumplimiento de la Gestión de Proyecto

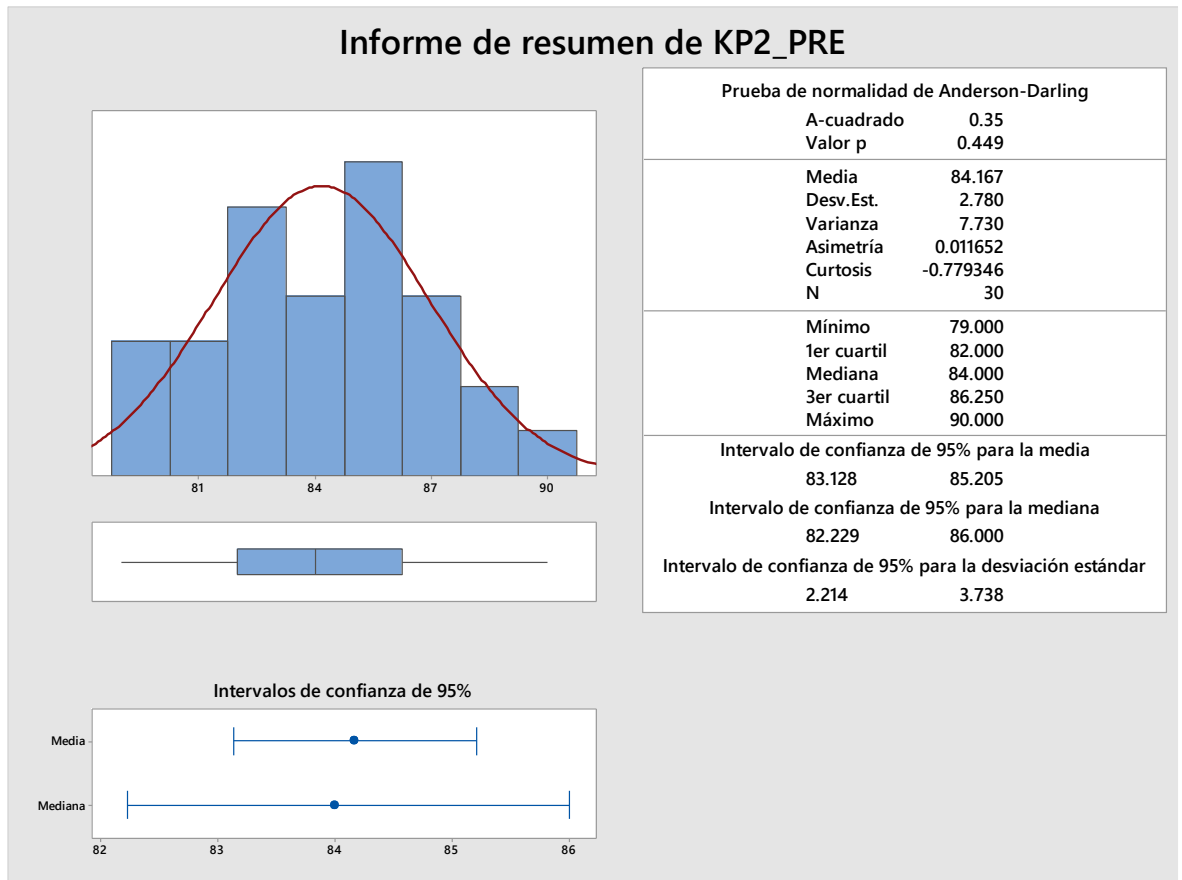
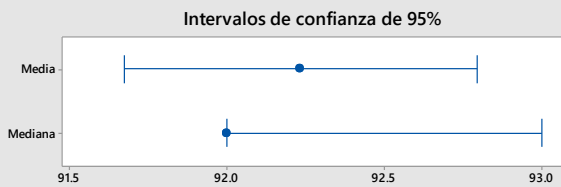
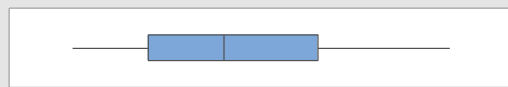
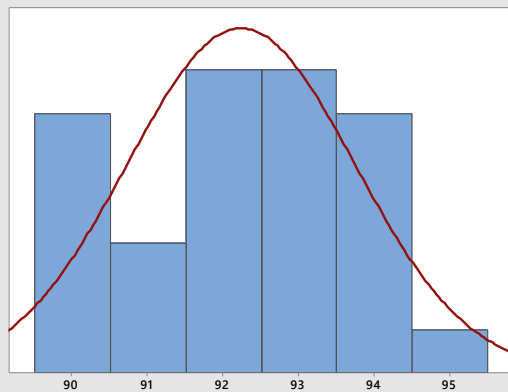


Figura 18. Estadística Descriptiva KPI2_PrePrueba

De la figura anterior se puede ver que los datos tienen una distribución normal puesto que el valor de $p > 0.05$ con una media de 84% y una desviación estándar de 2.7.

Informe de resumen de KP2_POST



Prueba de normalidad de Anderson-Darling

A-cuadrado 0.95
Valor p 0.014

Media 92.233
Desv.Est. 1.501
Varianza 2.254
Asimetría -0.16527
Curtosis -1.03553
N 30

Mínimo 90.000
1er cuartil 91.000
Mediana 92.000
3er cuartil 93.250
Máximo 95.000

Intervalo de confianza de 95% para la media

91.673 92.794

Intervalo de confianza de 95% para la mediana

92.000 93.000

Intervalo de confianza de 95% para la desviación estándar

1.196 2.018

Figura 19. Estadística Descriptiva KPI2_PostPrueba

De la figura anterior se puede ver que los datos tienen una distribución normal puesto que el valor de $p > 0.05$ con una media de 92% y una desviación estándar de 1.5.

C. KPI3: Porcentaje de cumplimiento en actividades de Desarrollo de MD

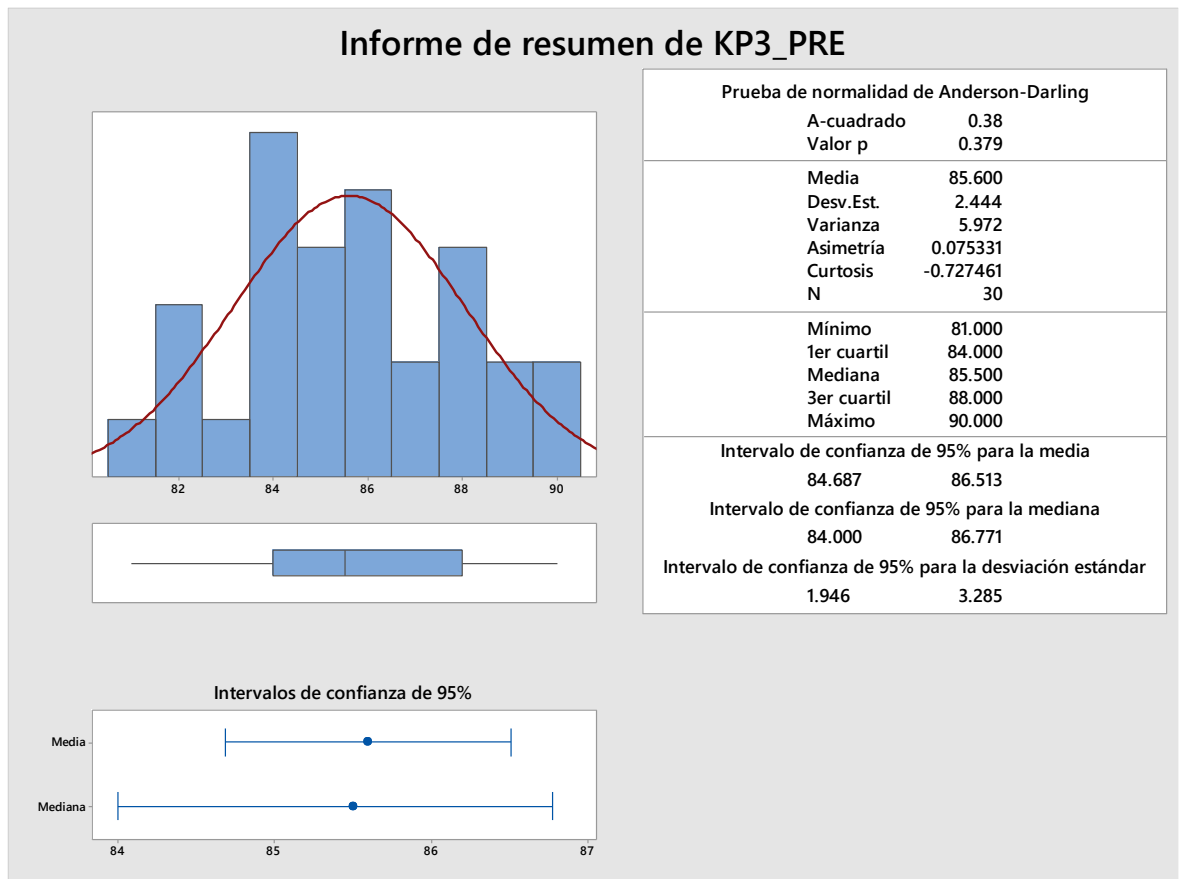


Figura 20. Estadística Descriptiva KPI3_PrePrueba

De la figura anterior se puede ver que los datos tienen una distribución normal puesto que el valor de $p > 0.05$ con una media de 85% y una desviación estándar de 2.4.

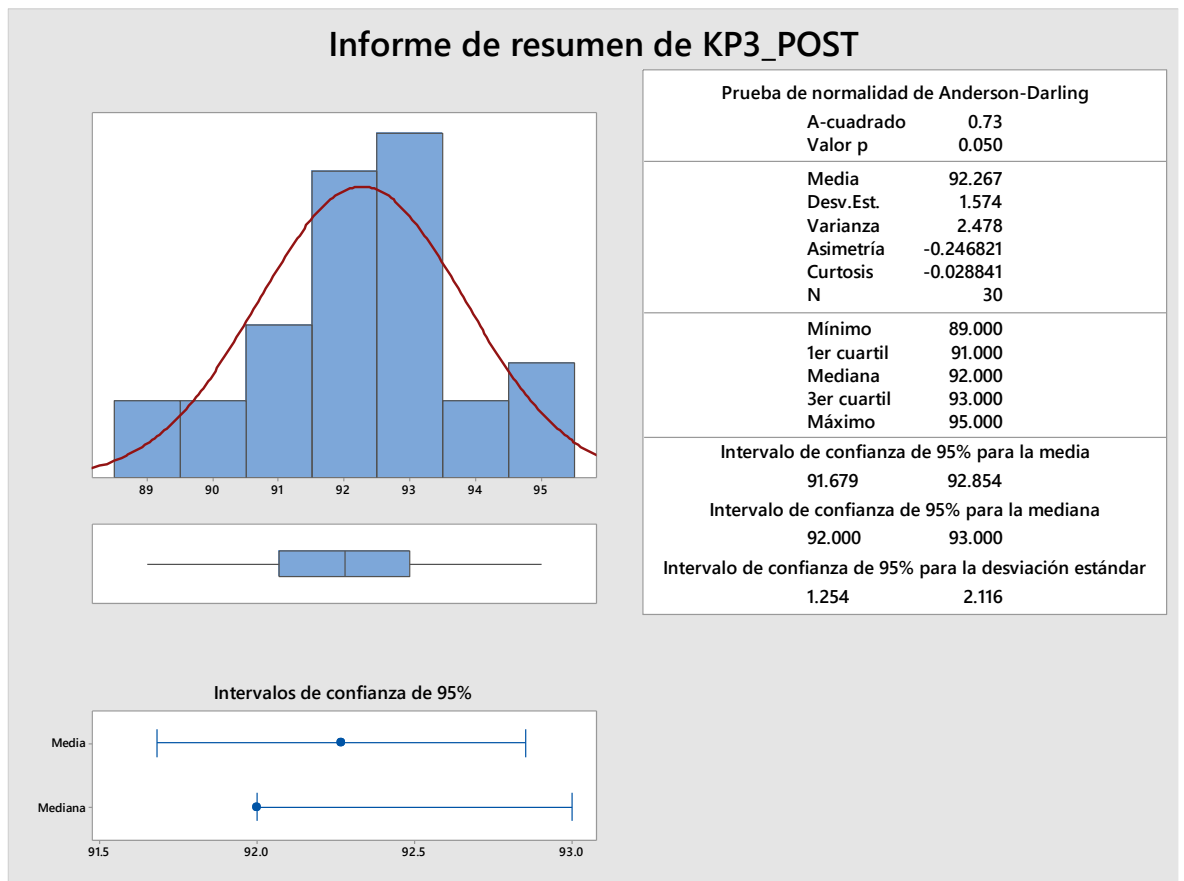


Figura 21. Estadística Descriptiva KPI3_PostPrueba

De la figura anterior se puede ver que los datos tienen una distribución normal puesto que el valor de $p > 0.05$ con una media de 92% y una desviación estándar de 1.5.

4.3. Contratación de Hipótesis

4.3.1. KPI1: Porcentaje de Cumplimiento Metodológico

a) Prueba de normalidad

Se puede apreciar en la Figura 15. del preprueba y la Figura 16. de la post prueba, los datos tienen una distribución normal, puesto que el valor de p es mayor a 0.05.

μ_1 = Media Porcentaje de Cumplimiento Metodológico en la Pre Prueba.

μ_2 = Media Porcentaje de Cumplimiento Metodológico en la Post Prueba

b) Planteamiento de Hipótesis

1. Hipótesis Alterna

La media de los Porcentaje de Cumplimiento Metodológico en la preprueba es menor al de la postprueba.

$$H_a: \mu_2 > \mu_1$$

2. b. Hipótesis Nula

H_0 . La media del Porcentaje de Cumplimiento Metodológico en la preprueba es igual al de la postprueba

$$H_0: \mu_2 = \mu_1$$

3. Nivel de significación: 5%

4. Estadístico de prueba: “t” de student

Prueba T e IC de dos muestras: KP1_PRE; KP1_POST

T de dos muestras para KP1_PRE vs. KP1_POST

	N	Media	Desv.Est.	Error estándar de la media
KP1_PRE	30	85.37	2.46	0.45
KP1_POST	88	94.12	2.70	0.29

$$\text{Diferencia} = \mu (\text{KP1_PRE}) - \mu (\text{KP1_POST})$$

Estimación de la diferencia: -8.756
Límite superior 95% de la diferencia: -7.864
Prueba T de diferencia = 0 (vs. <): Valor T = -16.43
Valor p = 0.000 GL = 54

5. Decisión

Como $p < 0,05$, se rechaza la H_0

6. Conclusión:

Los resultados de la prueba, aplicada porque los datos se distribuyen normalmente; demuestran que, como el resultado de p tiende a cero en relación a la probabilidad asumida de 0.05, se rechaza la hipótesis nula, y se acepta la hipótesis alterna.

4.3.2. KPI2: Porcentaje de Cumplimiento de la Gestión de Proyecto

a) Prueba de normalidad

Se puede apreciar en la Figura 17. del preprueba y la Figura 18. de la post prueba, los datos tienen una distribución normal, puesto que el valor de p es mayor a 0.05.

μ_1 = Media Porcentaje de Cumplimiento de la Gestión de Proyecto en la Pre Prueba.

μ_2 = Media Porcentaje de Cumplimiento de la Gestión de Proyecto en la Post Prueba

b) Planteamiento de Hipótesis

1. Hipótesis Alternativa

La media de los Porcentaje de Cumplimiento de la Gestión de Proyecto en la preprueba es menor al de la postprueba.

$$H_a: \mu_2 > \mu_1$$

2. b. Hipótesis Nula

H_0 . La media del Porcentaje de Cumplimiento de la Gestión de Proyecto en la preprueba es igual al de la postprueba

$$H_0: \mu_2 = \mu_1$$

3. Nivel de significación: 5%

4. Estadístico de prueba: “t” de student

Prueba T e IC de dos muestras: KP2_PRE; KP2_POST

T de dos muestras para KP2_PRE vs. KP2_POST

				Error estándar de la media
	N	Media	Desv.Est.	
KP2_PRE	30	84.17	2.78	0.51
KP2_POST	30	92.23	1.50	0.27

Diferencia = μ (KP2_PRE) - μ (KP2_POST)
Estimación de la diferencia: -8.067

Límite superior 95% de la diferencia: -7.097
Prueba T de diferencia = 0 (vs. <): Valor T = -13.98 Valor p =
0.000 GL = 44

5. Decisión

Como $p < 0,05$, se rechaza la H_0

6. Conclusión:

Los resultados de la prueba, aplicada porque los datos se distribuyen normalmente; demuestran que, como el resultado de p tiende a cero en relación a la probabilidad asumida de 0.05, se rechaza la hipótesis nula, y se acepta la hipótesis alterna.

Por consiguiente, el modelo aumenta el Porcentaje de Cumplimiento de la Gestión de Proyecto de manera significativa, mejorando el proceso la aplicación de minería de datos.

4.3.3. KPI3: Porcentaje de cumplimiento en actividades de Desarrollo de MD

a) Prueba de normalidad

Se puede apreciar en la Figura 19. del preprueba y la Figura 20. de la post prueba, los datos tienen una distribución normal, puesto que el valor de p es mayor a 0.05.

μ_1 = Media Porcentaje de cumplimiento en actividades de Desarrollo de MD en la Pre Prueba.

μ_2 = Media Porcentaje de cumplimiento en actividades de Desarrollo de MD Post Prueba

b) Planteamiento de Hipótesis

1. Hipótesis Alterna

La media de los Porcentaje de cumplimiento en actividades de Desarrollo de MD en la preprueba es menor al de la postprueba.

$$H_a: \mu_2 > \mu_1$$

2. b. Hipótesis Nula

Ho. La media del Porcentaje de cumplimiento en actividades de Desarrollo de MD en la preprueba es igual al de la postprueba

$$H_0: \mu_2 = \mu_1$$

3. Nivel de significación: 5%

4. Estadístico de prueba: “t” de student

Prueba T e IC de dos muestras: KP3_PRE; KP3_POST

T de dos muestras para KP3_PRE vs. KP3_POST

	N	Media	Desv.Est.	Error estándar de la media
KP3_PRE	30	85.60	2.44	0.45
KP3_POST	30	92.27	1.57	0.29

Diferencia = μ (KP3_PRE) - μ (KP3_POST)

Estimación de la diferencia: -6.667

Límite superior 95% de la diferencia: -5.777

Prueba T de diferencia = 0 (vs. <): Valor T = -12.56 Valor p = 0.000 GL = 49

5. Decisión

Como $p < 0,05$, se rechaza la Ho

6. Conclusión:

Los resultados de la prueba, aplicada porque los datos se distribuyen normalmente; demuestran que, como el resultado de p tiende a cero en relación a la probabilidad asumida de 0.05, se rechaza la hipótesis nula, y se acepta la hipótesis alterna.

Por consiguiente, el modelo aumenta el Porcentaje de cumplimiento en actividades de Desarrollo de MD de manera significativa, mejorando el proceso la aplicación de minería de datos.

**CAPÍTULO VI:
DISCUSIÓN DE
RESULTADOS**

6.1. Discusión

Cumplimiento Metodológico

El modelo propuesto en comparación con el modelo CRISP-DM ha mejorado el Cumplimiento Metodológico de manera significativa, mejorando el proceso la aplicación de minería de datos, como se puede apreciar en la siguiente tabla se evaluó 5 características respecto al modelo en sí, en ese sentido el modelo propuesto alcanzo un 94% en comparación con CRISP-DM que llegó a un 85%.

Tabla 15
Comparación de modelos a nivel de Modelo General

		CRISP-DM	MoPro-DM
DOMINIO	CARACTERÍSTICA		
Modelo	Definición de actividades específicas	85%	94%
	Especificación de pasos a seguir para cada actividad		
	Definición de las entradas de cada actividad (Input)		
	Definición de salidas de cada actividad (Output)		
	Entrega de una guía de buenas prácticas para cada una de las actividades.		

Cumplimiento de la Gestión de Proyecto

El modelo propuesto en comparación con el modelo CRISP-DM incluye actividades de la gestión de proyectos en sus diferentes etapas, lo que ha permitido mejorar significativamente la gestión del mismo, como se puede apreciar en la siguiente tabla donde el modelo tiene un 92% de cumplimiento en comparación con el modelo CRISP-DM que tiene un 84%

Tabla 16
Comparación de modelos a nivel del subproceso de Gestión de Proyectos

DOMINIO	SUBDOMINIO	CARACTERÍSTICA	CRISP-DM	MoPro-DM
Sub Proceso de Gestion de Proyectos	Aleance	Entregables del proyecto	84%	92%
		Actividades de control		
	Tiempo	Sucesión de actividades		
		Estimación de duración de cada actividad		
		Cronograma del proyecto		
	Costos	Afectación de recursos		
		Estimación de costos		
		Elaboración de presupuesto		
	RRHH	Planificación de RRHH		
		Interacción entre los miembros del equipo de trabajo		
		Medición de rendimiento.		
	Riesgo	Identificación de riesgos		
		Priorización de riesgos		
		Acciones correctivas y preventivas		

Cumplimiento en actividades de Desarrollo de MD

El modelo propuesto en comparación con el modelo CRISP-DM incluye actividades del subproceso de aplicación de minería de datos en sus diferentes etapas, lo que ha permitido mejorar significativamente su ejecución, como se puede apreciar en la siguiente tabla donde el modelo tiene un 92% de cumplimiento en comparación con el modelo CRISP-DM que tiene un 85%

Tabla 17
Comparación de modelos a nivel del subproceso de Minería

DOMINIO	SUBDOMINIO	CARACTERÍSTICA	CRISP-DM	MoPro-DM
Sub proceso de Minería	COMPRESIÓN DE LOS DATOS	Evaluación de la organización	85%	92%
		Identificación de stakeholders		
		Definición del problema		
		Evaluación de fuentes de datos		
		Análisis de soluciones potenciales		
		Definición de objetivos del proyecto		
		Determinación de criterio de éxito		
		Análisis de técnicas de DM		
		Especificación de documentos (entregables)		
	SELECCIÓN Y PREPARACIÓN DE LOS DATOS	Análisis Exploratorio		
		Limpieza de datos		
		Transformación de variables y creación de atributos derivados		
		Análisis descriptivo de datos depurados		
		Revisión del conjunto de datos final con el usuario		
	MODELADO	Selección de técnicas de Minería de Datos		
		Evaluación de resultados		
		Evaluación de nuevos modelos		
		Directivas para el descubrimiento de patrones		
	EVALUACIÓN	Interpretación de modelos		
		Comparación y ponderación de modelos		
		Revisión del proceso		
		Directivas para modelos no viables		
	IMPLEMENTACIÓN-DESPLIEGUE	Implementación de nuevos modelos		
		Programa de Mantenimiento		
		Resumen del proyecto		

CAPÍTULO VII: CONCLUSIONES

7.1. Conclusiones

- La industria de servicios es el sector más dinámico, se encuentra presente en todos los tipos de organizaciones y genera mayor porcentaje de ingresos, por tal razón las soluciones de Minería de Datos que permitan implementaciones ágiles y al mismo tiempo integren las necesidades del negocio son aquellas que permitan a los usuarios facilidad de comprensión y desarrollo de nuevas estrategias de modelos para la exploración de datos
- Existe una gran congruencia con la necesidad de las organizaciones en encontrar información relevante de acuerdo al comportamiento de sus clientes, en la industria de Servicios la problemática se ve más acentuada debido al nivel de demanda del usuario en función de los SLA con la organización, debido a esto en este trabajo se evidencia que las metodologías de Minería de datos plantean actividades específicas para mantener el control en todo el proceso en función con los objetivos de la organización y el Core del negocio.
- El modelo según las observaciones de los expertos mejoró en un 6% en el cumplimiento Metodológico a nivel de Fases.
- El modelo según las observaciones de los expertos mejoró en un 8% el Porcentaje de cumplimiento de la gestión del proceso de minería de datos.
- El modelo según las observaciones de los expertos mejoró en un 7% el porcentaje cumplimiento en actividades de Desarrollo de MD

CAPÍTULO VIII: RECOMENDACIONES

8.1. Recomendaciones

Realizar de manera minuciosa el desarrollo de las primeras fases del proceso de minería de datos, los datos de partida se consideran la base fundamental del ciclo, nunca se presenta el caso donde el volumen de información recolectado es exagerado o perjudicial para el análisis.

Las soluciones de Minería de Datos deben ser analizadas de acuerdo a la necesidad del negocio, se debe tomar en cuenta la infraestructura, recursos de la organización (volumen de datos, uso y precio) y, capacidad de integración, la decisión final debe basarse en el cumplimiento del objetivo del negocio dentro del alcance establecido al inicio del proceso.

La implementación de una herramienta de Minería de datos es responsabilidad de toda la organización. El enfoque debe ir más allá del conjunto de datos. Los aspectos de análisis para cada metodología deben ser ponderados de acuerdo al objetivo del proyecto y de la mano de la gestión de recursos, del entorno de aplicación y de las actividades de dirección de proyectos.

La selección de la herramienta no debe ser tomada en base al cumplimiento de una sola característica, aunque la misma sea la más importante. Es necesario validar todos los aspectos de cumplimiento que abarque el mayor número de características que permitan alcanzar el objetivo del proyecto y brinde mayor beneficio a la organización.

La implementación de la metodología en conjunto con la herramienta seleccionada debe procurarse con personal especializado que posea un conocimiento profundo del tema. La organización es responsable de que el personal que estará a cargo al finalizar el proyecto se integre durante la implementación y su capacitación constante en todo el proceso.

Las herramientas para Minería de Datos y su implementación suelen resultar en costos exorbitantes, un error de las empresas es considerarlos como gastos innecesarios (de acuerdo a la dimensión de la organización). Es recomendable considerar estos valores como inversiones a largo plazo, que en la línea de tiempo proporcionan valor al negocio con ventaja competitiva desplegando nuevas oportunidades.

CAPÍTULO IX: REFERENCIAS

- Arboleya, H. (2013). Propuesta de Ciclo de Vida y Mapa de Actividades para Proyectos de Explotación de Información. *Revista Latinoamericana de Ingeniería de Software*, 18.
- Baker, R. S. J. d, & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM / Journal of Educational Data Mining*, 1(1), 3–17. Recuperado de <https://jedm.educationaldatamining.org/index.php/JEDM/article/view/8>
- Baker, R. S. (2014). Educational Data Mining: An Advance for Intelligent Systems in Education. *IEEE Intelligent Systems*, 29(3), 78–82. <https://doi.org/10.1109/MIS.2014.42>
- Baker, R. S. J. (2011). Data Mining for Education. *International Encyclopedia of Education*, 7, 112–118. Recuperado de <http://www.columbia.edu/~rsb2162/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf>
- Ballesteros, Alejandro, Sánchez, Daniel, & García, Ricardo. (2013). Minería de datos educativa: Una herramienta para la investigación de patroness de aprendizaje sobre un contexto educativo. Recuperado el 27 de noviembre de 2018, de <http://web.b.ebscohost.com/abstract?site=ehost&scope=site&jrnl=18709095&AN=94988767&h=uZEy5J2ZZr9F15rq2mRER3tO5tZXCTpQmNrhGZhUgNzfl8yeXA%2bA%2fMe%2bHCW4k7RUHFIOkXkqz3BMenwigTSieA%3d%3d&crl=f&resultLocal=ErrCrlNoResults&resultNs=Ehost&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authtype%3dcrawler%26jrnl%3d18709095%26AN%3d94988767>
- Bichsel, J. (2012). Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations, 31.

- Campbell, J. P., DeBlois, P., & Oblinger, D. G. (2007). Academic Analytics: A New Tool for a New Era. Recuperado el 27 de noviembre de 2018, de <https://er.educause.edu/articles/2007/7/academic-analytics-a-new-tool-for-a-new-era>
- Cooper, A. (2012). A Brief History of Analytics, 21.
- Davenport, T. H. (2006). Competing on Talent Analytics. *harvard business review*. Recuperado de <http://cs.brown.edu/courses/cs295-11/competing.pdf>
- Davenport, Thomas H, Harris, J. G., & Morison, R. (2010). Analytics at Work: Smarter Decisions, Better Results, 6.
- Davenport, Thomas H, Harris, J., & Shapiro, J. (2010). Competing on Talent Analytics. *Harvard Business Review*, 7.
- Gallardo Arancibia, J. A. (2009). *Metodología para la definición de requisitos en proyectos de data mining* (phd). Facultad de Informática (UPM). Recuperado de <http://oa.upm.es/1946/>
- Ganesh, M., Han, E.-H. (Sam), Kumar, V., Shekhar, S., & Srivastava, J. (1996). *Visual Data Mining: Framework and Algorithm Development*.
- García, Enrique, Romero, Cristóbal, De Castro Lozano, Carlos, Calders, Toon, & Ventura, Sebastian. (2011). Association Rule Mining in Learning Management Systems. <http://dx.doi.org/10.1201/b10274-9>
- García-Martínez, R., Lelli, R., Merlino, H., Cornachia, L., Rodriguez, D., Pytel, P., & Arboleya, H. (2011). Ingeniería de proyectos de explotación de información para PyMEs. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/20017>
- Hofmann, M., & Tierney, B. (2009). Development of an Enhanced Generic Data Mining Life Cycle (DMLC), (18), 23.

- Huapaya, C. R., Lizarralde, F. A., Arona, G. M., & Massa, S. M. (2012). Minería de Datos Educativa en Ambientes Virtuales de Aprendizaje, 5.
- IBM. (2012). Manual CRISP-DM de IBM SPSS Modeler. Recuperado de <http://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- Igartua Lopez J. I, Errasti Lozares N, & Zabaleta Etxebarria, N. (2012). Análisis de la Relación Existente entre los Estándares de Gestión de Proyectos y los Factores Críticos para su Éxito.
 The Analysis of the Links between the Project Management Standards and. *6th International Conference on Industrial Engineering and Industrial Management*, 943–950. Recuperado de <http://adingor.es/congresos/web/articulo/detalle/a/2235>
- International Educational Data Mining Society. (2011). *Educational Data Mining*. Recuperado de <http://www.educationaldatamining.org/>
- Johnson, L., Adams, S, & Cummins, M. (2012). *The NMC Horizon Report: 2012 Higher Education Edition*. Recuperado de <https://www.nmc.org/pdf/2012-horizon-report-HE.pdf>
- Lee, S. W., & Kerschberg, L. (1998). A methodology and life cycle model for data mining and knowledge discovery in precision agriculture. En *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on* (Vol. 3, pp. 2882–2887). IEEE.
- Ljevo, Ž., & Vukomanović, M. (2014). Project Management Practiced in Public Project Stream of Bosnia and Herzegovina. *Procedia - Social and Behavioral Sciences*, 119, 692–701. <https://doi.org/10.1016/j.sbspro.2014.03.077>
- Marie Bienkowski, Mingyu Feng, & Barbara Means. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue*

- Brief*. U.S. Department of Education Office of Educational Technology.
 Recuperado de <https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf>
- Moine, J. M. (2013). *Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo* (Tesis). Facultad de Informática. Recuperado de <http://hdl.handle.net/10915/29582>
- Montequín, R., Teresa, M., Cabal, Á., Valeriano, J., Fernández, M., Manuel, J., & Valdés, G. (s/f). METODOLOGÍAS PARA LA REALIZACIÓN DE PROYECTOS DE DATA MINING. *DATA MINING*, 9.
- Moscoso-Zea, O., Andres-Sampedro, & Luján-Mora, S. (2016). Datawarehouse design for educational data mining. En *2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1–6).
<https://doi.org/10.1109/ITHET.2016.7760754>
- Mustaro, P., & Rossi, R. (2013). Project Management Principles Applied in Academic Research Projects. *Issues in Informing Science and Information Technology*, 10, 325–340. <https://doi.org/10.28945/1814>
- Nascimento, G., & Oliveira, A. (2013). AgileKDD: AN AGILE PROCESS MODEL TO KNOWLEDGE DISCOVERY IN DATABASES AND BUSINESS INTELLIGENCE SYSTEMATIZATION. Recuperado el 12 de junio de 2018, de https://updoc.site/download/agilekdd-contecsi_pdf
- NIAKŠU, O. (2015). Development and Application of Data Mining Methods in Medical Diagnostics and Healthcare Management, 176.
- Palacios, M. D. (2014). Inteligencia de Negocios para Empresas de Construcción y la Gestión de Proyectos con enfoque en las mejores prácticas. Recuperado de <http://dspace.uazuay.edu.ec/handle/datos/3739>

- Pérez, J. S. (2007). DEFINITION AND INSTANTIATION OF AN INTEGRATED DATA MINING PROCESS TIN2004-05873, 9.
- Project Management Institute, Inc. (2013). *GUÍA DE LOS FUNDAMENTOS PARA LA DIRECCIÓN DE PROYECTOS* (Quinta).
- Pyle Dorian. (2003). *Business Modeling and Data Mining*. Elsevier.
<https://doi.org/10.1016/B978-1-55860-653-1.X5000-3>
- Robin Way. (2013). Model Deployment: The Moment of Truth. Recuperado el 12 de junio de 2018, de <https://blogs.sas.com/content/sastraining/2013/10/14/model-deployment-the-moment-of-truth/>
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.
<https://doi.org/10.1016/j.eswa.2006.04.005>
- Sánchez, M., & César, J. (2017). Adaptación de estándares de dirección de proyectos particularizados para la minería de datos. Recuperado de <http://digibuo.uniovi.es/dspace/handle/10651/43633>
- SAS Institute. (2018). Data Mining and SEMMA :: Data Mining Using SAS(R) Enterprise Miner(TM): A Case Study Approach, Third Edition. Recuperado el 27 de noviembre de 2018, de <http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbj4n1xueveo2uoujy.htm>
- Solarte-Pazos, L., & Sánchez-Arias, L. F. (2014). Gerencia de proyectos y estrategia organizacional: el modelo de madurez en Gestión de Proyectos CP3M© V5.0. *Innovar*, 24(52), 5–18. <https://doi.org/10.15446/innovar.v24n52.42502>

Vanrell, J. Á., & Bertone, R. A. (2010). Modelo de proceso de operación para proyectos de explotación de información. Presentado en XVI Congreso Argentino de Ciencias de la Computación. Recuperado de <http://hdl.handle.net/10915/19328>

CAPÍTULO X:

ANEXOS

Anexo 01: SOLICITUD DIRIGIDA A EXPERTO

Sr. Ing. _____

Es grato dirigirme a Usted para manifestarle mi saludo cordial. Dada su experiencia profesional y méritos profesionales, le solicito su inapelable colaboración como experto para la validación del contenido de los ítems que conforma el instrumento (Anexo), que tiene como finalidad validar la investigación titulada: “MODELO DE PROCESO PARA LA APLICACIÓN DE MINERÍA DE DATOS EN LAS UNIVERSIDADES PERUANAS”, para obtener el grado académico de Doctor en Ingeniería de Sistemas.

Para efectuar la validación del instrumento, Usted deberá leer cuidadosamente cada enunciado y sus correspondientes alternativas de respuesta, en donde se pueda valorar de acuerdo a lo que ha podido verificar con el modelo.

Se agradece cualquier sugerencia respecto al contenido u otro aspecto que considere relevante para mejorar el mismo.

Atentamente.

Mg. José Luis Herrera Salazar

Anexo 2: Cuestionario para evaluar los modelos

			CRISP-DM	MoPro-DM
DOMINIOS	SUBDOMINIO	CARACTERÍSTICA		
MODELO	PROCESO	Definición de actividades específicas		
		Especificación de pasos a seguir para cada actividad		
		Definición de las entradas de cada actividad (Input)		
		Definición de salidas de cada actividad (Output)		
		Entrega de una guía de buenas prácticas para cada una de las actividades.		
Sub proceso de Minería	COMPRESIÓN DE LOS DATOS	Evaluación de la organización		
		Identificación de stakeholders		
		Definición del problema		
		Evaluación de fuentes de datos		
		Análisis de soluciones potenciales		
		Definición de objetivos del proyecto		
		Determinación de criterio de éxito		
		Análisis de técnicas de DM		
	Especificación de documentos (entregables)			
	SELECCIÓN Y PREPARACIÓN DE LOS DATOS	Análisis Exploratorio		
		Limpieza de datos		
		Transformación de variables y creación de atributos derivados		
		Análisis descriptivo de datos depurados		
		Revisión del conjunto de datos final con el usuario		
	MODELADO	Selección de técnicas de Minería de Datos		
		Evaluación de resultados		
		Evaluación de nuevos modelos		
		Directivas para el descubrimiento de patrones		
	EVALUACIÓN	Interpretación de modelos		
		Comparación y ponderación de modelos		
		Revisión del proceso		
		Directivas para modelos no viables		

	IMPLEMENTACIÓN- DESPLIEGUE		Implementación de nuevos modelos		
			Programa de Mantenimiento		
			Resumen del proyecto		
Sub Proceso de Gestion de Proyectos		Alcance	Entregables del proyecto		
			Actividades de control		
		Tiempo	Sucesión de actividades		
			Estimación de duración de cada actividad		
			Cronograma del proyecto		
		Costos	Afectación de recursos		
			Estimación de costos		
			Elaboración de presupuesto		
		RRHH	Planificación de RRHH		
			Interacción entre los miembros del equipo de trabajo		
			Medición de rendimiento.		
		Riesgo	Identificación de riesgos		
			Priorización de riesgos		
			Acciones correctivas y preventivas		