**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**

**Universidad del Perú, Decana de América**

**FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**

**ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS**

 

**SISTEMA BASADO EN MINERÍA DE DATOS PARA LA DETECCIÓN DE PATRONES COMERCIALES Y DE VENTA EN UNA EMPRESA COMERCIAL**

**Tesis para optar por el título profesional de:**

**INGENIERO DE SISTEMAS**

**Presentada por:**

**Santillán Rodríguez, Edgar Hugo**

Ciudad Universitaria, 28 de Abril de 2018

**Santillán Rodríguez, Edgar Hugo**

**SISTEMA BASADO EN MINERÍA DE DATOS PARA LA DETECCIÓN DE PATRONES COMERCIALES Y DE VENTA EN UNA EMPRESA COMERCIAL**

“Tesis presentada a la Universidad Nacional Mayor de San Marcos, Lima, Perú, para obtener el Título de Ingeniero de Sistemas”

Orientador: Vega Huerta Hugo Froilan

UNMSM – LIMA

Abril 2018

.

© Santillán Rodríguez, 2018.

Todos los derechos reservados.

Este trabajo está dedicado a mi familia por su constante apoyo hacia mi persona en todo momento hasta la actualidad.

**AGRADECIMIENTOS**

A mi tutor Vega Huerta Hugo Froilan, por su orientación, paciencia y confianza en la realización de esta obra.

A los profesores de la UNMSM, por ser modelos a seguir y parte importante en nuestra formación académica.

**SISTEMA BASADO EN MINERÍA DE DATOS PARA LA DETECCIÓN DE PATRONES COMERCIALES Y DE VENTA EN UNA EMPRESA COMERCIAL**

**RESUMEN**

El proyecto que se presenta en este documento tiene como objetivo la identificación de patrones comerciales y de venta en una Empresa comercial que podría predecir el comportamiento de la venta de algún producto por temporada; se realiza lo antes mencionado para que la Empresa comercial pueda tener un mayor conocimiento sobre las ventas que puede lograr a partir de información recopilada en los años anteriores, y con ello elaborar políticas que permitan sacar el máximo provecho al área comercial y pueda optimizar recursos en cierta medida.

Para lograr dichos objetivos se han de usar herramientas de software, herramientas de análisis de datos y metodologías que permitan realizar lo mencionado de acuerdo a lo esperado para ello se han de seleccionar herramientas estandarizadas y aceptadas internacionalmente en sus respectivos campos, en el caso de las herramientas de software, han de ser seleccionadas de acuerdo a la comparación de criterios y de acuerdo a los requerimientos que se necesiten , el mismo concepto aplica para las herramientas de análisis de datos.

En conclusión, el proyecto se lleva a cabo con éxito previniendo los efectos negativos o eventos inoportunos que puedan generarse durante su ejecución, además de mantener los objetivos generales y específicos con sus respectivos métodos para mantener la idea clara y concisa de lo que se pretende realizar desde el inicio de este.

**ÍNDICE**

[Capítulo 1: Introducción 7](#_Toc468233535)

[1.1 Antecedentes 7](#_Toc468233536)

[1.2 El problema 8](#_Toc468233537)

[1.2.1 Problema general. 8](#_Toc468233538)

[1.2.2 Problemas específicos. 8](#_Toc468233539)

[1.3 Objetivos 9](#_Toc468233540)

[1.3.1 Objetivo general 9](#_Toc468233541)

[1.3.2 Objetivos específicos 9](#_Toc468233542)

[1.4 Justificación 10](#_Toc468233543)

[Capítulo 2. Marco teórico 10](#_Toc468233544)

[2.1 Datos, información y conocimiento 11](#_Toc468233545)

[2.2 Data mining 11](#_Toc468233546)

[2.3 Knowledge Discovery in DataBases (KDD, Descubrimiento de Conocimiento en Bases de Datos) 13](#_Toc468233547)

[Capítulo 3. Estado del Arte 17](#_Toc468233548)

[Referencias bibliográficas 21](#_Toc468233549)

# Introducción

## Antecedentes

La mayoría de instituciones están comenzando a organizar, clasificar y usar correctamente la información, ya que conforme avanza la tecnología resulta más fácil almacenar grandes cantidades de datos, esta tendencia continua a un ritmo acelerado y permite obtener ventajas en campos tales como el marketing, criminalística, educación, y muchos más.

“La cantidad de información que nos llega cada día es tan inmensa que nos resulta difícil asimilarla”.1

En España, la minería de datos se aplica de forma exitosa en los procesos industriales, lo cual ha permitido agilizar el proceso en la producción del acero formando una red neuronal basada en la información obtenida.

“Los resultados obtenidos han servido para resaltar las principales ventajas que aportan las redes neuronales en el modelado de procesos industriales, versatilidad para adaptarse a diversas funciones, efectividad para modelar procesos no lineales aplicación intuitiva y facilidad de implementación”2

En algunas instituciones educativas de Colombia también se han logrado buenos resultados que permiten obtener nuevas posibilidades de su explotación, además de permitir desarrollar patrones generales de acuerdo al problema que se intenta solucionar.

 “Se ha obtenido un patrón general de deserción estudiantil en las dos IES determinado por un promedio bajo y el tener materias perdidas en los primeros semestres de la carrera. Se han determinado factores socioeconómicos y académicos asociados a la deserción estudiantil.” 3

En Perú, en el ámbito criminalístico se ha intentado desarrollar avances en cuanto al buen uso de la información, permitiendo implementar sistemas de tipo alerta que basados en la información de los datos permitieron hacer visible los conocimientos que tenían algunas comisarías de la ciudad de Chiclayo, lo que se tradujo en la reducción del tiempo y facilidad de acceso a sistemas informáticos colaborativos que pudieron hacer más eficiente su labor.

“Tuvo como hipótesis que el desarrollo de un modelo de minería de datos como herramienta de apoyo podría contribuir a la caracterización de perfiles delictivos.”4

##  El problema

### Problema general.

¿En qué medida la minería de datos permite detectar patrones comerciales y de ventas en una Empresa Comercial?

### Problemas específicos.

* ¿Qué metodología es la adecuada para poder implementar el Sistema de detección de patrones comerciales y de ventas en una Empresa Comercial mediante la minería de datos?
* ¿Cuáles son los procesos que usaremos para extraer la información requerida del Sistema de detección de patrones comerciales y de ventas en una Empresa Comercial?
* ¿Qué algoritmos usaremos para poder clasificar la información obtenida del Sistema de detección de patrones comerciales y de ventas en una Empresa Comercial?

##  Objetivos

### Objetivo general

Desarrollar un sistema basado en minería de datos para la detección de patrones comerciales y de ventas en una Empres Comercial.

### Objetivos específicos

* Elegir una metodología para implementar el Sistema de detección de patrones comerciales y de ventas en una Empresa Comercial
* Elegir un proceso para implementar el Sistema de detección de patrones comerciales y de ventas en una Empresa Comercial
* Implementar un algoritmo para el Sistema de detección de patrones comerciales y de ventas en una Empresa Comercial.

##  Justificación

“Estamos abrumados de datos y  hambrientos de conocimiento”5, se puede decir que con el proceso de globalización los datos ha comenzado no solo a mostrar información valiosa sino también a predecir fenómenos en distintas áreas, pero no basta con tener los datos, además de eso hay que entender y descifrar los patrones que guardan estos para poder sacarle un mejor provecho. Para lograr obtener un mejor beneficio y sacar provecho de la información comercial, es necesario conocer y explotar dicha data y generar una vista clara, por eso es necesario el sistema de detección de patrones comerciales y de ventas en una Empres Comercial para poder generar la toma de decisiones en las actividades futuras.

El desarrollo del sistema basado en minería de datos permitirá a la Empresa Comercial:

* Elaborar un historial de las ventas en base a cantidades registradas.
* Identificar el volumen producido, ordenarlos y clasificarlos.
* Elaborar políticas de control y desarrollo para poder calcular la cantidad de producción que puedan elaborar.

# Marco teórico

## Datos, información y conocimiento

Los datos son individualmente características, atributos o hechos sin ninguna información relevante. (WEISS & DAVIDSON, 2010). Estos no tienen mayor significado individualmente (dato), pero si su volumen es grande se puede interpretar como información. (WEISS & DAVIDSON, 2010)



Relación entre Datos, Información y Conocimiento (LIEW, 2007)

Una de las características de la información es que el volumen de información con respecto a los datos siempre es menor, sin embargo, la información tiene mayor valor para el negocio. (LIEW, 2007)

La información siempre tiene algún significado o interpretación que se puede sacar de la enorme cantidad de datos. (LIEW, 2007)

El resultado del procesamiento de información en conjunto con la experiencia humana es el conocimiento; este tiene mayor valor para el negocio, pero su volumen decrece con respecto a la información. (LIEW, 2007). Del conocimiento obtenido del pasado y presente se puede predecir futuros comportamientos del entorno. (LIEW, 2007)

## Data mining

La minería de datos es el proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones aparentemente caóticos que tienen una explicación que pueden descubrirse mediante diversas técnicas de esta herramienta.(Ángeles y Santillán, p. 79)

Es un proceso no trivial que tiene como entrada datos y como salida Información, en este proceso se hace un análisis detallado a través del uso de algoritmos para descubrir patrones o comportamientos de los datos. (WEISS & DAVIDSON, 2010)

Minería de Datos es un término genérico que engloba resultados de investigación, técnicas y herramientas usadas para extraer información útil de grandes bases de datos. (MOLINA & GARCIA, 2012)

Según (MOLINA & GARCIA, 2012) los procesos de Data mining tiene algunas tecnologías de apoyo como:

* Razonamiento Estadístico

Las técnicas y métodos estadísticos son muy usados para el Data Mining, debido a que juegan un papel importante en el análisis de datos, y también en el aprendizaje automático. Muchos paquetes estadísticos usados hoy en día se han integrado a diferentes bases de datos y se comercializan como productos de Data Mining. (MOLINA & GARCIA, 2012)

* Visualización

Las tecnologías de visualización muestran gráficamente los datos de la base de datos. Los modelos de visualización pueden ser bidimensionales, tridimensionales e incluso multidimensionales (en bases de datos). (MOLINA & GARCIA, 2012)

* Procesamiento paralelo

Se ha vuelto un factor crítico en base de datos y Data Mining (Minería de Datos), ya que el rendimiento de consultas es importante para la fluidez de procesos posteriores. (MOLINA & GARCIA, 2012)

* Apoyo a la toma de decisiones

Son básicamente herramientas que se usan para tomar decisiones eficaces y se basan en teoría de decisiones. Las herramientas de apoyo de toma de decisiones se usan para eliminar los resultados innecesarios obtenidos del proceso de Data Mining. (MOLINA & GARCIA, 2012)

* Aprendizaje automático

Consiste en aprender las experiencias del pasado con respecto a alguna métrica de rendimiento. Por ejemplo en los videojuegos se aprende de las experiencias pasadas para aprender a jugar, En Data Mining se usan muchas de las técnicas de aprendizaje automático en los algoritmos. (MOLINA & GARCIA, 2012)

## Knowledge Discovery in DataBases (KDD, Descubrimiento de Conocimiento en Bases de Datos)

“Proceso no trivial de identificar patrones válidos, novedosos potencialmente útiles y, en última instancia, comprensibles a partir de los datos”. (Fayad, Piateski-Shapiro y Smyth, 1996, citado en Camargo y Silva, 2010 p. 12)

Según UIAF (2014), el proceso de Knowledge Discovery in DataBases (KDD) surte cuatro pasos para la generación de conocimiento. Estas etapas pueden ser recursivas, es decir, que se retorna a ellas una y otra vez (proceso iterativo) a medida que se obtienen resultados preliminares que requieren replantear las variables iniciales.



Proceso KDD (UIAF, 2014)

* **Selección de los datos**

Es la etapa más dispendiosa (requiere cerca del 90% del tiempo), como se describe en Fayyad et. al. (1996), ya que consiste en la recolección y preparación de los datos. En este proceso se comprende la problemática asociada a la base de datos y se establecen objetivos. A la vez, se identifican las variables que serán consideradas para la construcción del modelo de minería de datos (MD). De acuerdo con el origen de cada tabla(s) de datos, se establece el mecanismo que se utilizará para extraer la información requerida por el modelo; también incide la opinión del experto, quien puede sugerir las series que explican parcialmente la variable objetivo (clase)

* **Pre procesamiento de datos**

**Integración de datos:** Se analiza si la base de datos requiere incluir o integrar información o variables que reposan en otras bases de datos, y que será relevante para el modelo de minería de datos. Si es necesario, se realiza un modelo de entidad-relación entre tablas, el cual permite representar las entidades relevantes (representaciones gráficas y lingüísticas) de un sistema, así como sus propiedades e interrelaciones (Ávila, 2005).

**Reconocimiento y limpieza:** Se depura el conjunto de datos respecto a valores atípicos, faltantes y erróneos (eliminación de ruido e inconsistencias).

* **Selección de características**

**Exploración y limpieza de datos:** Aplicando técnicas de análisis exploratorio de datos (estadístico, gráfico, entre otros), se busca identificar la distribución de los datos, simetría, pruebas de normalidad y correlaciones existentes entre los datos. En esta etapa es útil el análisis descriptivo del conjunto de datos (clustering y segmentación, escalamiento, reglas de asociación y dependencia, reducción de la dimensión), identificación de datos nulos, ruido y outliers, así como el uso de matrices de correlación (si las variables son numéricas), diagramas (barras, histogramas, caja y bigotes), entre otras técnicas adecuadas de muestreo.

**Transformación:** Se estandariza o normaliza la información (colocarla en los mismos términos de formato y forma). La selección de la técnica a aplicar dependerá del algoritmo que se utilizará para la generación de conocimiento. Las técnicas comúnmente utilizadas son: discretización, escalado (simple y multidimensional) y estandarización.

**Reducción de datos:** Se disminuye el tamaño de los datos mediante la eliminación de características redundantes. Selección/extracción de atributos: Se realiza un proceso de identificación y selección de variables relevantes. Entre las técnicas más utilizadas para este proceso se encuentran, métodos basados en filtros y en wrappers (seleccionan los atributos en función de la calidad del modelo de MD asociado a los atributos utilizados). Construcción de atributos: En caso que los datos se encuentren asociados a patrones complejos, se construye un atributo sencillo que facilite la interpretación del algoritmo. Se puede recurrir a diferentes técnicas, como la construcción guiada por los datos, el modelo o el conocimiento; finaliza cuando se han analizado y seleccionado las técnicas de extracción de conocimiento que mejor se adapten al lote de datos.

* **Minería de Datos**

La minería de datos según Esteban (2008) et. al. (1991 / 1995), se puede definir como un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos, que a su vez, facilita la toma de decisiones y emplea técnicas de aprendizaje supervisado y no-supervisado. En la fase se define el tipo de investigación a realizar de acuerdo con las características de los atributos que conforman la base de datos, que pueden ser de tipo descriptivo4 y/o predictivo. El siguiente paso es la elección y aplicación de métodos de extracción de conocimiento y la selección de medidas de validación para evaluar el rendimiento y la precisión del(os) método(s) escogido(s) por el analista de MD.

A continuación se presentan las técnicas empleadas, las cuales pueden ser de tipo supervisado o no supervisado.

**Identificación:** Evidenciar la existencia de objetos, eventos y actividades en el conjunto de datos (análisis factorial, discriminante, regresivo, de correlaciones).

**Clasificación:** Particionar los datos de acuerdo a las clases o etiquetas asignadas al conjunto de datos (ej: tablas de decisión, reglas de decisión, clasificadores basados en casos, redes neuronales, clasificadores bayesianos y clasificadores basados en acoplamientos).

**Agrupación:** Permitir la maximización de similitudes y minimización de diferencias entre objetos, mediante la aplicación de algún criterio de agrupación.

**Asociación:** Tener presente que las reglas de asociación buscan descubrir conexiones existentes entre objetos identificados.

**Predicción:** Descubrir el comportamiento de ciertos atributos en el futuro. (Regresión y series temporales, análisis discriminante, métodos bayesianos, algoritmos genéricos, árboles de decisión, redes neuronales).

El proceso más importante en minería de datos es el proceso de limpieza de los datos y definición de las variables. Si los datos no son correctos el modelo creado no servirá. Del mismo modo, la validez de los patrones descubiertos depende de cómo se apliquen al mundo real o a las circunstancias. Aunque integra el conocimiento de diferentes fuentes como la Biología, Matemáticas, Estadística, Ciencias Sociales, entre otras, la minería de datos presenta diferencias importantes, por ejemplo, respecto a la estadística, la cual plantea una hipótesis y usa datos para probarla o refutarla. La eficacia de este enfoque se ve limitada por la creatividad del usuario para desarrollar las diversas hipótesis, así como por la estructura del software que utiliza; por el contrario, la minería de datos emplea una aproximación de descubrimiento para examinar simultáneamente varias relaciones en bases de datos multidimensionales, identificando aquellas que se presentan con frecuencia (2008).

* **Interpretación y Resultados**

Se analizan los resultados de los patrones obtenidos en la fase de MD, mediante técnicas de visualización y de representación, con el fin de generar conocimiento que aporte mayor valor a los datos. En esta fase se evalúan los resultados con los expertos y, si es necesario, se retorna a las fases anteriores para una nueva iteración.

# Estado del Arte

**LA MINERÍA DE DATOS COMO UN MÉTODO INNOVADOR PARA LA DETECCIÓN DE PATRONES DE DESERCIÓN ESTUDIANTIL EN PROGRAMAS DE PREGRADO EN INSTITUCIONES DE EDUCACIÓN SUPERIOR (2013)**

Ricardo Timarán Pereira, Andrés Calderón Romero

Universidad de Nariño

San Juan de Pasto, Colombia

**Resumen**

En Latinoamérica, la educación superior presenta altas tasas de deserción estudiantil, especialmente en los primeros semestres académicos, conllevando 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. En este artículo se presenta uno de los resultados del proyecto de investigación financiado por el Ministerio de Educación Nacional cuyo objetivo fue detectar patrones de deserción estudiantil a partir de los datos socioeconómicos, académicos, disciplinares e institucionales de los estudiantes de los programas de pregrado de la Universidad de Nariño e Institución Universitaria CESMAG, dos instituciones de educación superior (la primera pública y la segunda privada) de la ciudad de Pasto (Colombia), utilizando técnicas de Minería de Datos. Inicialmente, se seleccionaron, de las bases de datos de estas instituciones, los datos socio-económicos, académicos, disciplinarios e institucionales de los estudiantes que ingresaron en los años 2004, 2005 y 2006 a los diferentes programas de pregrado, con el fin de hacerles un seguimiento hasta el año 2011, para determinar la deserción de los estudiantes. Se construyó un repositorio de datos utilizando el SGBD PostgreSQL en el cual se aplicaron técnicas para obtener datos limpios a los cuales se les aplicò las técnicas de minería de datos. Obteniendo al final un resultado favorable los cuales sirvieron para definir políticas para la retención estudiantil.

Metodología : Descubrimiento De Conocimiento En Bases De Datos (KDD).

Técnica : La técnica de clasificación basada en árboles de decisión.

Algoritmo : Algoritmo de árboles de decisión C4.5.b (r)

Conclusiones y trabajos futuros: Se ha obtenido un patrón general de deserción estudiantil en las dos IES determinado por un promedio bajo y el tener materias perdidas en los primeros semestres de la carrera. Se han determinado factores socioeconómicos y académicos asociados a la deserción estudiantil. La evaluación, análisis y utilidad de estos patrones permitirá soportar la toma de decisiones eficaces de las directivas universitarias enfocadas a formular políticas y estrategias relacionadas con los programas de retención estudiantil que actualmente se encuentran establecidos. Como trabajos futuros están el continuar con el estudio de deserción estudiantil en la Universidad de Nariño e Institución Universitaria CESMAG aplicando otras técnicas de minería de datos tales como asociación y clustering con el fin de determinar afinidades, similitudes y relaciones entre los factores socioeconómicos y académicos de las estudiantes que desertan.

**Conclusión**

Se ha logrado encontrar muchas causantes de deserción estudiantil, uno de los más notorios de los indicadores encontrados es el promedio bajo que obtienen los estudiantes durante los primeros semestres de la carrera, también se logró encontrar otro indicador muy marcado, los factores socioeconómicos, con esto se ha logrado obtener un patrón general de deserción estudiantil a través de la minería de datos, que luego de ser clasificada y tratada, va a brindar un panorama más detallado que deberán de ser tomadas en cuenta por las autoridades respectivas. Gracias a la minería de datos se ha logrado un avance en cuanto a la organización y buen uso de la información en Colombia, como se observa en esta publicación, si los datos son analizados, seleccionados y filtrados correctamente pueden ser una fuente muy rica de conocimiento, lo que va a permitir realizar acciones correctivas y preventivas en dichas instituciones de Colombia.

**IDENTIFICACIÓN Y DETECCION DE PATRONES DELICTIVOS**

**BASADA EN MINERIA DE DATOS. (2007)**

Perversi, I.1, Valenga, F.2, Fernández, E.3, 4, Britos P.3, 4, García-Martínez, R.3, 4

1 Departamento de Ingeniería Industrial. ITBA

2 Facultad de Informática Ciencias de la Comunicación y Técnicas Especiales. UM

3 Centro de Ingeniería de Software e Ingeniería del Conocimiento. Escuela de Postgrado. ITBA

4 Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA

**Resumen**

En este artículo se describen los resultados preliminares de la línea de investigación sobre el uso de minería de datos aplicados a la identificación y detección de patrones delictivos, analizando los homicidios dolosos cometidos en la República Argentina.

En la actualidad el SAT (Sistema de Alerta Temprana) se encuentra implementado en todas las provincias reportando información de los hechos delictivos ocurridos en todo el país. Esta información está siendo tratada a través de análisis estadístico sin hacer uso de técnicas ni herramientas de minería de datos. Por lo tanto no se obtiene un verdadero aprovechamiento de los datos obtenidos para lo cual el planteamiento de la aplicación de la minería de datos en el proceso de identificación y detección de patrones delictivos comenzando con el análisis de homicidios dolosos cometidos en la República Argentina, es factible.

El problema se abordó desde el principio con clusterizar los datos relevantes homicidios dolosos cometidos para luego pasar a analizar los cluster obtenidos y validarlos con los usuarios y por último aplicar algoritmos de inducción a cada cluster para encontrar explicaciones descriptivas del comportamiento que subyace a la pertenencia a los mismos.

Obteniendo así conocimiento sobre el dominio no identificable mediante otros métodos.

Estrategia : Proceso de Clustering, análisis y validación de los clústeres, aplicar algoritmos de inducción a cada cluster.

Técnica : La técnica de descripción basada en Clustering y la técnica de clasificación basada en arboles de decisión.

Algoritmo : Algoritmo C4.5 implementado en el Algoritmo J48, algoritmo K-means.

Conclusiones: Existe información a partir de la cual es posible desarrollar un proyecto de Minería de Datos a gran escala para ayudar a la generación de políticas criminales en la República Argentina. Los conocimientos descubiertos como resultado de este proceso sirven para: [a] proporcionar una justificación sustentada en los datos disponibles de los conceptos preexistentes y [b] la detección de piezas de conocimiento sobre el dominio no identificable mediante otros métodos. Se propone continuar con el proyecto: [a] aplicando técnicas de inducción para explicar en mayor detalle los cluster identificados y [b] ampliar el análisis a otros ámbitos y tipos de hechos (por ejemplo: homicidios dolosos causados por accidentes de tránsito).

**Conclusión**

Aprovechando la gran base de datos a la cual tienen acceso, y usando técnicas implementadas en Java como J48 para tratar dicha información, se ha logrado obtener piezas de conocimiento sobre el dominio no identificable, y se han logrado agruparlos en clusters (donde los elementos agrupados tienen características en común) los cuales permitirán ayudar a la elaboración de políticas criminales en Argentina.

# Referencias bibliográficas

* **Artículos**

1 Luis Carlos Molina Félix (2002). *Data mining: torturando a los datos hasta que confiesen.* Consultado en Noviembre de 2016. Recuperado de: <http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>

3 Jiménez J. (2013). “*La Minería de Datos como un Método innovador para la detección de Patrones de deserción estudiantil en Programas de Pregrado en instituciones de educación superior”.* Universidad de Nariño, San juan de Pasto, Colombia. Consultado en Noviembre de 2016. Recuperado de: <http://www.acofipapers.org/index.php/acofipapers/2013/paper/viewFile/211/112>

Perversi, I., Valenga, F., Fernández, E., Britos P., García-Martínez, R (2007). *“Identificación y detección de patrones delictivos basada en Minería de Datos”.* Departamento de Ingeniería Industrial. ITBA, Facultad de Informática Ciencias de la Comunicación y Técnicas Especiales. UM, Centro de Ingeniería de Software e Ingeniería del Conocimiento. Escuela de Postgrado. ITBA, Laboratorio de Sistemas Inteligentes. Facultad de Ingeniería. UBA, Argentina. Consultado en Noviembre de 2016. Recuperado de:

<http://iidia.com.ar/rgm/comunicaciones/WICC-07-385-389.pdf>

Vallejos S. (2006). “*Trabajo de adscripción de Minería de Datos”.* Universidad Nacional del Nordeste Facultad de Ciencias Exactas, Naturales y Agrimensura. Consultado en Noviembre de 2016. Recuperado de: <http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf>

Minería de Datos (Data Mining).  *¿Qué es y para qué sirve?* *(2ª parte)* *(DV00106A).* Consultado en Noviembre de 2016. Recuperado de: <http://www.aprenderaprogramar.com/index.php?option=com_attachments&task=download&id=203>

Trabajo de adscripción de Minería de Datos. *Universidad Nacional del Nordeste Facultad de Ciencias Exactas, Naturales y Agrimensura,* Sofia J. Vallejos (online). Consultado en Noviembre de 2016. Recuperado de: <http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf>

Liew, A. (2007). Understanding data, Information, Knowledge and their Inter-Relationships. *Journal of Knowledge Management Practice.* Vol. (8), 11-14 Consultado en Noviembre de 2016. Recuperado de:

<http://www.tlainc.com/articl134.htm>

Molina, J. y García, J. (2012). Técnicas de análisis de datos 5-12 Consultado en Noviembre de 2016. Recuperado de:

<http://ocw.uc3m.es/ingenieria-informatica/analisis-de-datos/libroDataMiningv5.pdf>

Weiss, G. y Davidson, B. (2010*).Data Mining. The Handbook of technology Management*, Vol (3), 1-17. Consultado en Noviembre de 2016. Recuperado de:

<http://storm.cis.fordham.edu/gweiss/papers/data-mining-chapter-2010.pdf>

UIAF (2014). *Técnicas de minería de datos para la detección y prevención del lavado de activos y la financiación del terrorismo (LA/FT),* 14-18. Consultado en Noviembre de 2016. Recuperado de:

<http://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Tecnicas-de-mineria-de-datos-para-la-prevencion-de.pdf>

Camargo, H. y Silva, M. (2010). Dos caminos en la búsqueda de patrones por medio de Minería de Datos: SEMMA y CRISP. *Revista de tecnología* Vol. (9), 11-12. Consultado en Noviembre de 2016. Recuperado de:

<http://www.uelbosque.edu.co/sites/default/files/publicaciones/revistas/revista_tecnologia/volumen9_numero1/dos_caminos9-1.pdf>

Ángeles, M. y Santillán, A. (2008). *Minería de Datos, Conceptos, características, estructura y aplicaciones*. Consultado en Noviembre de 2016. Recuperado de:

<http://www.ejournal.unam.mx/rca/190/RCA19007.pdf>

Flórez, R. y Fernández, J. (2008). *Las Redes Neuronales Artificiales, Fundamentos teóricos y aplicaciones*. Consultado en Noviembre de 2016. Recuperado de:

<https://books.google.com.pe/books?id=X0uLwi1Ap4QC&pg=PA7&lpg=PA7&dq=redEs+neuronales&source=bl&ots=gMOElqpm_h&sig=uZ5tF1tQIg5qvFOc4Rll78TwWG8&hl=es419&sa=X&ved=0ahUKEwjkj4DGu53QAhVJySYKHSaPAnE4ChDoAQhDMAY#v=onepage&q=redes%20neuronales&f=false>

WorPress.com. *Curso de Analítica Predictiva y Minería de Textos con RapidMiner y*

*CRISP-DM*. Consultado en Noviembre de 2016. Recuperado de:

[https://addkw.com/e-learning/formacion-en-analitica-avanzada/curso-de-analitica- predictiva-y-mineria-de-textos-con-rapidminer-y-crips-dm/](https://addkw.com/e-learning/formacion-en-analitica-avanzada/curso-de-analitica-%20predictiva-y-mineria-de-textos-con-rapidminer-y-crips-dm/)

* **Tesis**

2 Gonzales A. (2007). *“Desarrollo de Técnicas de Minería de Datos en procesos industriales”* (Tesis doctoral). Universidad de la Rioja, La Rioja, España. Consultado en Noviembre de 2016. Recuperado de: <https://dialnet.unirioja.es/servlet/tesis?codigo=1166>

4 Jimenez L. (2015). “*Aplicación de un Sistema de alerta temprana basada en la Minería de Datos para identificar patrones delictivos en la Ciudad de Chiclayo”.* (Tesis para optar el título de Ingeniero de Sistemas y Computación). Universidad Católica Santo Toribio de Mogrovejo, Chiclayo, Perú. Consultado en Noviembre de 2016. Recuperado de: <http://tesis.usat.edu.pe/jspui/handle/123456789/583>

* **Páginas web**

UOC (2002). Consultado en Noviembre de 2016. Recuperado de:

<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>

5 DATA MINING CONSULTING. *¿Para qué sirve la minería de datos?* Consultado en Noviembre de 2016. Recuperado de:

<http://www.dataminingperu.com/blog_dmc/13-blog/63-para-que-sirve-la-mineria-de-datos>