

El uso de modelos y metodologías de minería de datos para la inteligencia de negocios

CENDEJAS-VALDEZ, José Luis¹†, ACUÑA-LÓPEZ, Miguel Ángel¹, CORTES-MORALES, Griselda², BOLAÑOS-JIMÉNEZ, Gerardo³

¹Universidad Tecnológica de Morelia, CA - PRODEP Transferencia tecnológica para la construcción de software.

²Universidad Autónoma de Coahuila – Unidad Norte, CA - PRODEP Tecnologías Computacionales.

³Universidad Tecnológica de Morelia- Tecnologías de la Información y Comunicación.

Recibido Mayo 29, 2017; Aceptado Junio 06, 2017

Resumen

El descubrimiento de conocimiento a partir de grandes cantidades de información almacenada en bases de datos, brinda a las organizaciones una ventaja competitiva y estratégica; por lo que la adopción y explotación de la minería de datos representa un reto en infraestructura tecnológica, en almacenamiento y procesamiento de la información. Existen diversas herramientas comerciales y de código abierto como lo son, On-line Analytical Processing (OLAP), XLStat, Weka, Statistical Package for the Social Sciences (SPSS) y Orange que tienen como fundamento modelos de referencia, algoritmos y metodologías para la explotación de datos entre ellos destacan Cross Industry Standard Process of data mining (CRISP-DM), Sample, Explore, Modify, Model, and Assess (SEMMA) y Catalyst los cuales brinda información referente a la recopilación, uso y manejo de datos. El análisis de los datos a través de las herramientas y/o metodologías antes mencionadas da como resultado la definición de patrones y relaciones que existen entre los datos recopilados y analizados, brindando un apoyo para la generación de inteligencia de negocios; lo que permite a las organizaciones contar con fundamentos para la correcta y oportuna toma de decisiones. En la presente investigación se realizó un comparativo de los modelos/metodologías, así como las herramientas de software más sobresalientes, tomando como fundamento el modelo CRISP-DM y a los algoritmos que sustentan la propuesta metodológica de esta investigación.

Modelos de minería de datos, software de minería de datos, aplicaciones de código abierto e inteligencia de negocios

Abstract

The discovery of knowledge from large amounts of information stored in databases, gives organizations a competitive and strategic advantage; so the adoption and exploitation of datamining represents a challenge in technological infrastructure, storage and processing of information. There are several commercial tools and open source as they are On-line Analytical Processing (OLAP), XLStat, Weka, Statistical Package for the Social Sciences (SPSS) and Orange, which are based on reference models, algorithms and methodologies for the exploitation of data, among them stand out Cross Industry Standard Process of data mining (CRISP-DM), Sample, Explore, Modify, Model, and Assess (SEMMA) and Catalyst, which provide information regarding the collection, use and handling of data. The analysis of the data through the aforementioned tools and / or methodologies results in the definition of patterns and relationships that exist between the collected and analyzed data, providing support for the generation of business intelligence; Which allows organizations to have foundations for the correct and timely decision-making. In this research was made a comparison of the models / methodologies, as well as the most outstanding software tools based on the model CRISP-DM and the algorithms that support the methodological proposal of this research.

Datamining models, datamining software, open source software and business intelligence

Citación: CENDEJAS-VALDEZ, José Luis, ACUÑA-LÓPEZ, Miguel Ángel, CORTES-MORALES, Griselda, BOLAÑOS-JIMÉNEZ, Gerardo. El uso de modelos y metodologías de minería de datos para la inteligencia de negocios Revista de Sistemas Computacionales y TIC's. 2017, 3-8: 61-67.

† Investigador contribuyendo como primer autor.

Introducción

En la actualidad las organizaciones identifican a uno de los activos más importantes que les permite lograr una ventaja estratégica y competitiva la “*información*”; por lo que resulta imperante la necesidad de poder conocer e identificar esta información de una manera estructurada que se pueda convertir en datos. El identificar estos datos conlleva a valorar y utilizar procesos nuevos como la minería de datos que es una de las principales ramas de las tecnologías de la información.

Para las organizaciones, el aprovechamiento, desarrollo e implementación de la minería de datos representa un reto tanto en infraestructura tecnológica como en almacenamiento y procesamiento de la información, también se debe de considerar y valorar que la captación y capacitación de personal que pueda adaptar e innovar este tipo de tecnología es una instancia relevante que se debe de considerar.

La minería de datos es un conjunto de técnicas y herramientas de software que permiten extraer información de grandes cantidades de datos al identificar y reconocer una serie de patrones y algoritmos como los utilizados y ligados con la inteligencia artificial.

Esta técnica es considerada viable ya que se fundamenta en cuatro vertientes de la tecnología como lo son: 1) recolección masiva de datos, 2) alta capacidad de procesamiento de información, 3) algoritmos de minería de datos y 4) entrenamiento de sistemas orientados a la inteligencia de negocios.

Existen diversos tipos de herramientas como lo son las encaminadas al software libre que resulta más adecuado con un costo más bajo que el software comercial que genera ventajas importantes como ser más fiable, robusto, seguro y de costo casi nulo.

La inteligencia de negocios o business intelligence (BI), permite manejar e interpretar cúmulos de información, a través de un conjunto de procesos, algoritmos matemáticos, aplicaciones y tecnologías para obtener información rápida, sencilla y precisa en el análisis de datos, brindando con ello la solución de problemáticas y soporte en la toma de decisiones, trascendental hoy en día para las organizaciones.

Con ello fundamenten el logro de sus objetivos y metas a través de la inteligencia de negocios generada por la minería de datos contando así con un nivel de madurez digital elevado debido a que dichas instancias se sustentan a través de la información documentada durante los últimos ejercicios llevados a cabo en la generación de productos y/o servicios de las organizaciones.

Por lo anterior las organizaciones tienen la capacidad de fundamentar sus propios modelos de negocio a través de las correlaciones de las diversas variables, las cuales sustentan con la información esencial de la generación de productos y/o servicios de calidad, esto permitirá que las organizaciones sean punta de lanza en su ramo. Por lo que el objetivo general de esta investigación es generar el diseño de un modelo que brinde las etapas mínimas necesarias para la aplicación de algoritmos matemáticos con la finalidad de brindar a las organizaciones los elementos necesarios para generar la inteligencia de negocios en las organizaciones de nuestro país.

Marco Teórico

Existen diversas metodologías y técnicas que nos ayudan en la recopilación y análisis de datos, por lo que en primera instancia se describirá una de las bases más importante para la realización del actual documento.

¿Qué es la minería de datos?; quizás la definición más significativa de la minería de datos tiene como origen el “Descubrimiento de Conocimiento a partir de Bases de Datos”, donde se define como el “proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos” (Silva & Camargo, 2010).

Con la definición anterior se puede establecer la finalidad de la minería de datos con la cual se trata de descubrir patrones, anomalías y tendencias a través del análisis de la información, no se puede confiar totalmente en los resultados obtenidos por una aplicación y garantizar con ello el éxito y rentabilidad del estudio, se debe de vincular y trasladar los datos a la toma de decisiones basadas en los resultados obtenidos. SAS Instituto especifica el concepto de minería de datos, como el proceso de seleccionar, explorar, modificar, modelizar y valorar grandes cantidades de datos con el objetivo de descubrir patrones desconocidos que puedan ser utilizados como ventaja competitiva respecto a los competidores. (Pérez & Santín, 2007)

La minería de datos o descubrimiento de conocimiento en bases de datos, es una herramienta tecnológica eficaz de grandes dimensiones en la búsqueda y selección de información inédita y potencialmente útil a partir de un gran cumulo de datos de información. La minería de datos, define patrones y las relaciones que existen en los datos analizados, automatizando este proceso y proporciona resultados que pueden ser utilizados en un sistema de apoyo para la toma de decisiones estratégicas de la organización.

La búsqueda de patrones requiere que los datos aun no procesados sean simplificados de manera sistemática, con la finalidad de no abarcar la información de forma específica y dar prioridad a datos más generales.

Para con ello extraer e identificar los eventos aislados que apoyen en la generación de dichos patrones. Para llevar a cabo este proceso se utilizan varios algoritmos, que procesan los datos y apoyan para su localización basados en estándares validados.

La minería de datos es un concepto que necesita ser percibido a través de diferentes etapas, las cuales son: 1) la recopilación de información, 2) análisis, 3) predicciones, 4) la oportuna toma de decisiones, 5) entrenamiento en sistemas de información y 6) desarrollo de la inteligencia de negocios. Estos aplicativos han dado pauta a la creación de herramientas que permiten la recopilación y manejo de datos para el desarrollo de las etapas mencionadas con anterioridad.

Es necesario tomar como referencia modelos y/o metodologías que permitan a las organizaciones desarrollar un proceso eficiente en las predicciones basadas en históricos reales generados por ellas mismas, con el objetivo de dar cumplimiento a las metas de la organización; buscando con ello brindar un nivel de competitividad y un crecimiento económico de la región, en la tabla 1 se presentan diversos métodos que sirven para la ejecución e implementación en la minería de datos.

Nombre	Predictivo			Descriptivo	
	Clasificación	Regresión	Agrupamiento	Reglas de asociación	Correlaciones/ Factorizaciones
Redes Neuronales	✓	✓	✓		
Árboles de decisión ID3	✓				
Árboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓	✓	✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓			✓
Regresión logística	✓			✓	
Kmeans			✓		
A priori				✓	
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis factorial					✓
TwoStep, Cobweb			✓	✓	
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores soporte	✓	✓	✓		
CN2 (rules)	✓			✓	
Análisis discriminante	✓				
multivariante					
CRISP-DM	✓	✓	✓	✓	✓

Tabla 1. Métodos y algoritmos que sustentan la minería de datos.

Fuente: Hernández J., 2004.

Se debe considerar que un sistema de información es un conjunto de herramientas asociadas que proporcionan acceso a las actividades de gestión y se especializa en brindar información sobre cambios rápidos y oportunos; la información que solicita dicho programa es en gran parte de tipo numérica - cuantitativa y suele ser representada gráficamente. Este tipo de sistemas son los que dan vida a la minería de datos a través de estándares que son unidades o cadenas de información que tienen estructuras de frecuencia notoria. Dentro de estos estándares se puede hacer mención de algunas metodologías implementadas. Entre las más importantes se encuentran:

A) CRISP-DM sugerida por SPSS, la cual no solo garantiza una adecuada planeación sino una mayor efectividad en los resultados de un proyecto de minería de datos. (Chapman, 2007). B) SEMMA, la cual detalla un proceso más que un conjunto de herramientas analíticas; donde a partir de una muestra estadísticamente representativa de los datos, donde es más fácil aplicar técnicas estadísticas de exploración y visualización, seleccionar y transformar las variables más significativas e importantes para predecir los resultados y confirmar la exactitud de un modelo. (Zapata, 2010).

Estas metodologías se encuentran complementadas a través de diversas herramientas (software), que apoyan de forma eficaz y eficiente el llevar a cabo las acciones necesarias para implementar la minería de datos en diversos ámbitos.

Los modelos y metodologías son un conjunto de reglas, fórmulas y ecuaciones que pueden utilizarse para predecir un resultado basándose en un conjunto de campos o variables de entrada. Estos modelos requieren de herramientas especializadas (software) que se encuentren fundamentados en algoritmos matemáticos para el buen manejo de la selección de los datos obtenidos.

Los cuales se clasifican con base en su naturaleza: a) comerciales y b) software libre. Las herramientas tecnológicas, como cualquier otra herramienta, están diseñadas para facilitar el trabajo y permitir que los recursos sean aplicados eficientemente intercambiando información y conocimiento dentro y fuera de las organizaciones. (Universidad Metropolitana, 2013).

Existe una amplia gama de herramientas de software relacionadas con la gestión de datos, dentro de las cuales podemos encontrar algunas como:

1. Herramientas OLAP, son menos puntuales y funcionan sobre un sistema de información el cual permite ejecutar agregaciones y combinaciones de los datos de manera más compleja, con objetivos de análisis más estratégicos. Las herramientas OLAP están fundamentadas, en interfaces multidimensionales. Con la cual se proporciona facilidad para manejar y transformar los datos, lo que conlleva a producir otros datos más específicos para la oportuna toma de decisiones. (Núñez, 2011).
2. XLStat es un software estadístico de fácil uso para Microsoft Excel, utiliza técnicas pioneras de cálculo, de modo que permite resultados viables a velocidades imbatibles: utilizando cómputo paralelo, también cuenta con métodos de aprendizaje de máquinas (reglas de asociación, árboles de regresión y clasificación y K-Nearest Neighbors) y regresión de mínimos cuadrados parciales. (Addinsoft, 2015).

3. SAS desarrolla un ciclo de vida analítico interactivo paso a paso en el proceso de los datos para la toma de una decisión. SAS utiliza componentes integrados para reducir el tiempo para obtener valor del ciclo de vida del modelado eliminando los pasos redundantes y apoyando la cohesión en la cadena de información de los datos a la gestión de decisiones. Los procesos consistentes y las tecnologías para el desarrollo e implementación de modelos reducen los riesgos involucrados en el proceso de modelado al tiempo de apoyar la colaboración y el gobierno entre los miembros clave del negocio y de Tecnologías de la Información. (SAS, 2015).
4. SPSS Clementine es una herramienta integrada de minería de datos que incluye diversas fuentes de datos (ASCII, XLS, ODBC, etc.), un interfaz visual basado en procesos/flujo de datos, distintas herramientas de minería de datos (correlación, reglas de asociación, regresión, segmentación, clasificación, redes neuronales, reglas y árboles de decisión, etc.), manipulación de datos (pick & mix, muestreo, combinación y separación, etc.) y combinación de modelos. Las muestras de SPSS utilizan una complejidad que permite seleccionar una muestra acorde con un diseño específico y complejo capaz de incorporar los detalles del diseño del análisis de los datos para asegurar la eficacia de los resultados obtenidos. (Pérez & Santín, 2007).
5. Weka es un software libre disponible bajo General Public License (GNU) el aplicativo se caracteriza por tener una colección de herramientas de visualización y algoritmos para el análisis de datos y el modelado predictivo, junto con interfaces gráficas de usuario para un acceso fácil a esta funcionalidad. Esta serie de algoritmos de aprendizaje automático es bastante versátil para resolver problemas de minería de datos en el mundo real. (Sudhir, 2013).
6. Orange es una herramienta de minería de datos para el análisis de datos de forma interactiva basado en componentes de procedimientos de minería de datos. Su uso principal es la exploración de datos en la cual que se probaron y anotaron diferentes combinaciones de algoritmos de pre-procesamiento y aprendizaje mediante validación cruzada. (Zupan, 2012).

En la tabla 2 se presentan algunas de las herramientas más significativas y utilizadas en la minería de datos.

Aplicaciones de minería de datos	Creador/Desarrollador
OLAP	Microsoft
XLStat	Addinsoft
SAS	SAS Institute
IBM SPSS	IBM
Weka	Universidad de Waikato
Orange	Facultad de informática de la Universidad de Ljubljana

Tabla 2 Herramientas de minería de datos

Fuente: Elaboración propia

Metodología

Para llevar a cabo la investigación se realizó un estudio que está fundamentado en siete etapas, las cuales incluyen desde:

propuesta de la investigación, en la cual se plasman los objetivos y las preguntas de investigación con el objetivo de identificar el alcance y las metas a cubrir.

Factibilidad de la investigación, etapa donde se identificaron las relaciones entre las variables directas e indirectas y se determinó si era factible el llevar a cabo la investigación planteada.

Análisis de técnicas y/o modelos de minería de datos, algoritmos matemáticos, entrevistas con expertos, en esta parte de la investigación se abordó el análisis de las diferentes propuestas y herramientas que se encuentran inmersas en la explotación de la minería de datos así como los puntos de vista de expertos en la generación de la inteligencia de negocios de diferentes organizaciones.

Preparación de la información a través del modelo para la minería de datos denominado CRISP, donde se plasmó el proceso a seguir en la preparación de los datos.

La aplicación de algoritmos matemáticos los que permiten llevar a cabo diferentes pruebas estadísticas a la información recabada a través del paso anterior.

Predicciones, es el medio por el cual se obtienen los resultados.

Generación de la inteligencia de negocios, la cual permite dotar a las organizaciones de los fundamentos para la toma de decisiones con elementos estadísticos y con un gran porcentaje de certeza para alcanzar los objetivos planteados.

Estas etapas conforman al modelo metodológico propuesto en dicha investigación, el cual se muestra en la imagen 1.

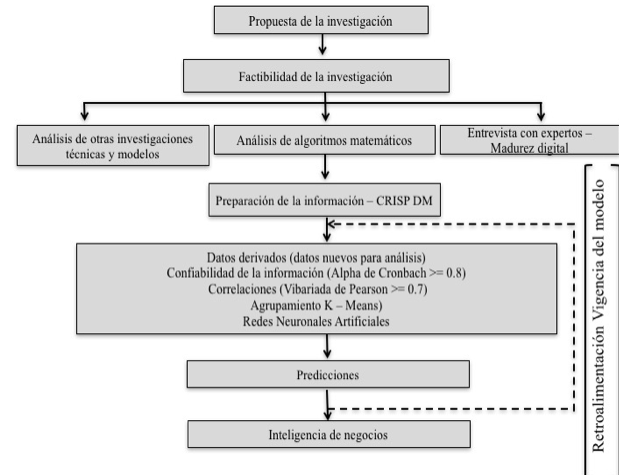


Figura 1 Modelo metodológico para el uso de minería de datos.

Fuente: Elaboración propia

El determinar un modelo metodológico ayudo a identificar la naturaleza de la investigación la cual es de tipo: cuantitativa, explicativa, de campo y transeccional (Kothari, 2004).

Se seleccionó la investigación cuantitativa y explicativa debido al alcance de la investigación: “determinar la relación entre el uso de modelos y metodologías más usadas en la minería de datos”. Además, de elegir la investigación de campo debido a que la fuente de información es con relación a responsables del área de TI y que han tenido alguna experiencia con herramientas y modelos en el uso de minería de datos. Se considera transeccional debido a que no se realizaron comparaciones históricas en un período de tiempo. (Vega, 2008)

Para el desarrollo de la investigación se prosiguió a la evaluación de las herramientas de software especializado en el uso de los diferentes algoritmos matemáticos que aportan a la generación de la minería de datos y por ende en la toma de decisiones para la inteligencia de negocios.

Las aplicaciones analizadas en este caso son **1) OLAP, 2) XLStat, 3) SAS, 4) SPSS de IBM, 5) Weka y 6) Orange** y que soportan el trabajo llevado a cabo bajo los siguientes algoritmos **a) Confiabilidad** (Alpha de Cronbach), **b) Correlaciones** (bivariada de Pearson), **c) Agrupamiento** (K – means), **d) Redes neuronales artificiales (RNA)** y con base a los resultados generados de dichos algoritmos permite fundamentar la inteligencia de negocios de una organización. Como se muestra en la tabla 3.

Algoritmos	Aplicaciones de minería de datos					
	OLAP	XLStat	SAS	IBM SPSS	Weka	Orange
Análisis de datos	✓	✓	✓	✓	✓	✓
Confiabilidad	✓			✓		
Correlaciones	✓		✓	✓		
Agrupamiento					✓	✓
Redes neuronales artificiales	✓			✓		✓
Predicciones	✓		✓	✓	✓	✓

Tabla 3 Relación de aplicación que soportan algoritmos para la minería de datos

Fuente: Elaboración propia

Resultados

El primer elemento de la propuesta metodológica hace referencia a:

Análisis de datos, esta es la parte fundamental de la propuesta debido a que se debe de gestionar y conformar la información recabada por años en las organizaciones a través de surveys y bases de datos bien organizadas, respetando así un modelo relacional con un filtrado en la información almacenada. Esto permite tener la seguridad de la información recopilada, la que sirve para la aplicación de las diferentes pruebas que fundamentan la toma de decisiones.

El estudio de **Alpha de Cronbach**, ayuda a generar el nivel de certeza y el nivel de confiabilidad de la información recopilada y analizada. La premisa de ello es que debe de cumplir con un valor igual o mayor a 70, este estudio sirve para contar y determinar que la información analizada no se repite, es verídica y confiable.

Dicho resultado se aprecia en la imagen 2. De lo contrario en caso de no cumplir con este indicador podrá existir una regresión al análisis de los datos quien funge como primer elemento.

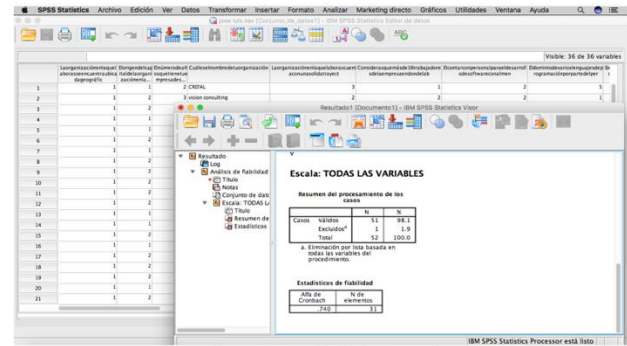


Figura 2. Estudio de confiabilidad con SPSS de IBM

Fuente: Elaboración propia

3) El estudio de Correlaciones, permite conocer las relaciones más fuertes y perfectas entre dos o más variables, estas representan a los ítems almacenados en las bases de datos y permite también conocer las relaciones más débiles, las cuales ayudan a identificar y conocer aspectos en los cuales hay que poner atención y que pueden servir para mejorar. Un ejemplo de ello se puede generar con SPSS de IBM, tal cual como se muestra en la imagen 3.

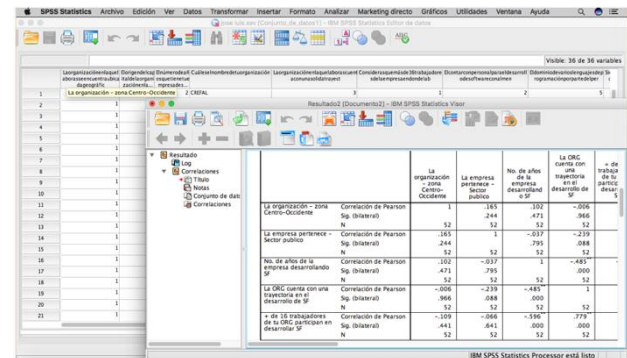


Figura 3 Estudio de correlaciones con SPSS de IBM

Fuente: Elaboración propia

4) El estudio de agrupamiento a través de conglomerados o cluster (K - means), este permite segmentar la información acumulada en grandes bases de datos tanto relacionales como conformadas en un datamart; este análisis busca que los elementos de cada cluster sean similares (homogeneidad entre los elementos). El nombre que se da es debido a que los centros de los clusters son las medias (K-means) de las observaciones asignadas a cada grupo además de que permite graficar y conocer los diferentes aglomerados y las características de cada uno de ellos y de los elementos que los conforman, como se muestra en la imagen 4.

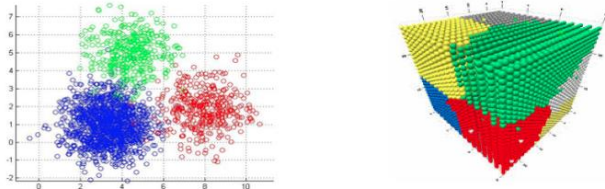


Figura 4 Estudio de agrupamiento K - means
Fuente: *Elaboración propia*

5) El estudio de redes neuronales artificiales (RNA) permite definir a un conjunto de elementos de procesamiento altamente interconectados y que son capaces de aprender con la información que ellos alimentan.

La principal característica de las RNA es que se aplica a un número de problemas complejos y heredan tres características básicas: a) paralelismo masivo, b) respuesta no lineal y c) procesamiento de información a través de múltiples capas, como se muestra en la imagen 5.

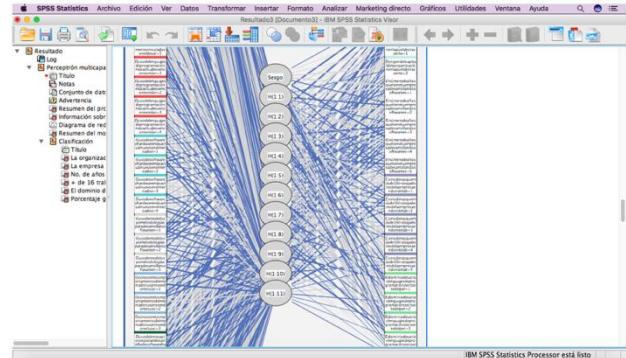


Figura 5 Estudio de RNA con SPSS de IBM
Fuente: *Elaboración propia*

El análisis e implementación de algoritmos con SPSS de IBM permite determinar la mejor opción para conjuntar una serie de pruebas para que cualquier organización pueda aplicarlas a través de un conjunto de datos históricos almacenados lo que permitirá generar la inteligencia de negocios para su correcta toma de decisiones.

La propuesta metodológica que se realiza para la generación de la inteligencia de negocios se encuentra conformada por seis estudios los cuales que se ven reflejados en las siguientes etapas: **1)** análisis de datos, **2)** confiabilidad – Alpha de Cronbach, **3)** correlaciones – bivariada de Pearson, **4)** agrupamiento – K_means, **5)** redes neuronales artificiales – RNA y esto conlleva a las **6)** predicciones. Este proceso se muestra en la imagen 6 y permite dotar a la organización de la implementación de sus propias herramientas en la toma de decisiones.

La propuesta metodológica sugerida en la imagen 6 se está implementando en la empresa Frutas de la Montaña, dedicada a generar mermeladas de blueberry en nuestro país con el objetivo de conocer las características inherentes de los clientes consumidores de dicho producto. (www.frutasdlamontana.mx)

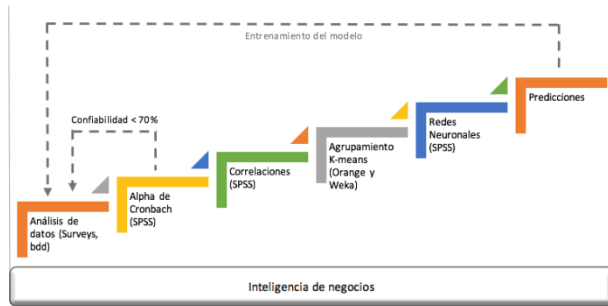


Figura 6 Propuesta metodológica para generar inteligencia de negocios en las organizaciones.

Fuente: Elaboración propia

Conclusiones

La toma de decisiones se ha vuelto un arma primordial en las organizaciones en la actualidad; la recopilación, procesamiento, uso y manejo de una gran cantidad de información brindan a ellas una ventaja competitiva en el sector en el cual se desarrollan. La toma de decisiones debe tener como finalidad lograr que se cumplan los objetivos y metas planteadas por la organización a través de su planeación estratégica.

Existen pocos estudios que dan a conocer las herramientas y/o modelos que permiten la implementación de la minería de datos, por lo que en esta investigación se presenta un conjunto de etapas que conforman a una propuesta metodológica, la cual permite generar la inteligencia de negocios en las organizaciones. Generando así un proceso que aporta de manera significativa e integral la generación de la madurez digital; ya que para ello es necesario tomar en cuenta la información, procesos, recurso humano y uso de herramientas tecnológicas.

Dicho con un sustento a través de diversas etapas, como lo son: **1)** la recopilación y análisis de datos, **2)** nivel de confiabilidad a través del alpha de Cronbach, **3)** estudio de correlaciones, **4)** clustering a través de K-means, **5)** generación de redes neuronales artificiales y **6)** generación de predicciones.

De igual manera se pretende dar a conocer un análisis comparativo que presenta las ventajas y desventajas de los modelos y/o metodologías de minería de datos, mostrando así los diversos escenarios donde la minería de datos tiene por objetivo colaborar en la solución de las problemáticas de las organizaciones, mediante la serie de etapas que fundamentan a la inteligencia del negocio.

Con base en el estudio comparativo generado en la presente investigación se determinó que SPSS de IBM fundamenta y colabora en todas las etapas de la propuesta metodológica planteada generando así un análisis predictivo y confiable, además de generar soluciones inmediatas e interactivas de su rendimiento actual con la capacidad de predecir resultados futuros y actuar proactivamente para el cumplimiento de los mismos.

Referencias

- Addinsoft. (2015). *XLSTAT su solución de análisis de datos*. Obtenido de <https://www.xlstat.com/es/>
- Camargo Hermandó, S. M. (2010). Dos caminos en la búsqueda de patrones por medio de minería de datos: SEMMA y CRISP. *Revista de tecnología*, 11-18.
- Chapman, C. K. (2007). *CRISP-DM 1.0: Step-by-step data mining guide*. Academic Press.
- Hernández J., M. J. (2004). *Introducción a la minería de datos*. Madrid, España: Pearson Educación, S.A.
- Kothari, C. (2004). *Research Methodology Methods and Techniques*. Jaipur, India: New Age International (P) Limited, Publishers.

Metropolitana, U. (2013). *Derecho a la educación*. Obtenido de Herramientas tecnológicas: <https://sites.google.com/site/grupo6derechoalaeducacion/educacion-universitaria/herramientas-tecnologicas>

http://ailab.ijs.si/dunja/TuringSLAIS-2012/Papers/Demsar_Orange.pdf

Núñez Cárdenas Felipe de Jesús, U. A.-H. (2011). *Introducción a la Minería de Datos*. Obtenido de https://www.uaeh.edu.mx/docencia/P_Presentaciones/huejutla/sistemas/mineria_datos/mineria.pdf

Pérez López Cesar, S. G. (2007). *Minería de datos Técnicas y Herramientas*. Madrid, España: Thomson.

SAS. (2015). *Gestionando el ciclo de vida analítico para las decisiones a escala*. Obtenido de http://www.sas.com/content/dam/SAS/es_mx/doc/assets/27-gestionando-ciclo-vida-2.pdf

Sudhir B. Jagtap, D. K. (2013). Census Data Mining Data Analysis using WEKA. *International Conference in "Emerging Trends in Science, Technology and Management-2013"*, 35-40.

Vega Escobar, A. M. (2008). Sistemas de Información como una Estrategia de Desarrollo, Crecimiento y Competitividad, de las Pequeñas y Medianas Empresas del Sector Servicios de Vigilancia en Bogotá D. C. *Avances en Sistemas e Informática.*, 121-130.

Zapata, G. A. (2010). *Minería de datos*. Obtenido de La minería de datos como herramienta para la toma de decisiones estratégicas: <http://gustavovalencia.net/app/webroot/img/Documents/BI/Actividades/001/Articulo%20DM.pdf>

Zupan, J. D. (2012). *ORANGE: data mining fruitful and fun*. Obtenido de