

## **TÍTULO: APLICACIÓN WEB QUE UTILIZA MINERÍA DE DATOS PARA GENERAR CONOCIMIENTO.**

### **Autores:**

**Sonia Benavides García**, Ingeniera Industrial, Master en Ciencias, Profesora Auxiliar de la Universidad de Matanzas “Camilo Cienfuegos”,  
[sonia.benavides@umcc.cu](mailto:sonia.benavides@umcc.cu)

**Erwin Hernández García**, Licenciado en Matemáticas y Computación, Máster en Ciencias, Analista informático de CUBARON SA, [erwin@fujae.var.cyt.cu](mailto:erwin@fujae.var.cyt.cu)

**Efraín Sánchez Pérez**, Licenciado en Matemáticas, Doctor en Ciencias Técnicas, Profesor Titular de la Universidad de Matanzas “Camilo Cienfuegos”,  
[efraín@fujae.var.cyt.cu](mailto:efraín@fujae.var.cyt.cu)

**Juan Pedro Febles Rodríguez**, Licenciado en Matemáticas, Doctor en Ciencias Técnicas, Profesor Titular, Director de Bioinfo, [febles@bioinfo.cu](mailto:febles@bioinfo.cu)

### **Resumen**

En la actualidad el descubrimiento de bases de datos (KDD en inglés) está siendo empleado como una variante novedosa, alcanzando excelentes resultados sobre grandes volúmenes de información. Este trabajo presenta una aplicación informática que utiliza técnicas de Minería de Datos, que permitirá generar conocimiento, ayudando así a la toma de decisiones de los directivos de la empresa CUBARON. La aplicación además presenta una interfaz con el soft libre WEKA para todo el tratamiento de minado.

### **Introducción**

El conocimiento es un recurso estratégico para el desarrollo económico y social contemporáneo. La información es el elemento básico principal en el proceso de adquisición, generación, gestión y transmisión del conocimiento. Las tecnologías, métodos y herramientas asociadas con estos procesos se han desarrollado notablemente en los últimos años. La aparición de Internet ha facilitado compartir, en puntos distantes, los resultados científicos.

El aumento continuo de la disponibilidad de datos, en particular, a partir de las redes de comunicaciones y la aplicación de la computación de alto desempeño, convierten en imprescindible el empleo de técnicas y herramientas que le den sentido y utilidad a la información existente.

En los últimos años del presente siglo, ha alcanzado un auge la minería de datos, soporte de filosofías como la gestión de las relaciones de una organización con sus clientes. Su fin es explorar y analizar las bases de datos disponibles para ayudar a la toma de decisiones en las organizaciones.

El surgimiento de técnicas como la minería de datos está asociado con la necesidad de procesar y analizar grandes volúmenes de datos, a fin de obtener información – mediante la consolidación de los datos- y conocimientos, útiles a la toma de decisiones por parte del personal directivo de una organización.

Por otra parte la fase de minería de datos tiene como objetivo construir modelos basados en los datos recopilados y debidamente preparados. Estos modelos obtenidos pueden ser empleados para realizar predicciones (*modelos predictivos*) o para entender mejor los datos (*modelos descriptivos*).

El presente trabajo presenta una aplicación útil que sirve de herramienta para la toma de decisiones gerenciales que permite la explotación de la información para la extracción del conocimiento.

### **Desarrollo:**

La aplicación que presentamos ha sido desarrollada utilizando Java Server Pages (JSP), que es una tecnología basada en el lenguaje de programación Java, que permite la implementación de sitios Web dinámicos. JSP fue desarrollado por Sun Microsystems con el objetivo de crear una plataforma para la programación en el lado del servidor, por lo tanto se puede definir que son páginas que contienen código HTML, script de cliente y un script que se ejecuta en el servidor, dando como resultado código HTML.

JSP ofrece las siguientes ventajas:

- Multiplataforma.
- Reusabilidad de componentes mediante el uso de Javabeans.
- Utilización de un lenguaje orientado a objetos como Java que ofrece grandes posibilidades.

El gestor de bases de datos MySQL fue el utilizado para el mantenimiento del mismo, debido a sus amplias posibilidades, rapidez, capacidad de almacenamiento de datos y fácil transportabilidad hacia diferentes Sistemas Operativos, incluyendo además posibilidades para su administración a través de la red.

La aplicación se apoya en un sistema desarrollado por la Universidad de Waikato en Nueva Zelanda para realizar los análisis de minería de datos, esta herramienta recibe el nombre de WEKA, acrónimo de *Waikato Environment for Knowledge Analysis*, WEKA se distribuye como software de libre distribución y código abierto desarrollado en Java.

Existen dos niveles en los que podemos usar el WEKA:

- Como usuario, o sea, para analizar datos contenidos en nuestras empresas, aplicándoles los algoritmos que el sistema tiene implementado usando la interfaz creada por sus desarrolladores
- Desarrollando nuestras propias aplicaciones que hagan uso de las librerías de clases del sistema y así crearnos nuestra propia interfaz.( en este nivel se enmarcó nuestro trabajo )

El WEKA se centra principalmente en los algoritmos de clasificación y filtrado, aunque cuenta también con algoritmos para extraer reglas de asociación y separación por clusters.

Nuestra aplicación se centra en el empleo de técnicas de Minería de Datos como son:

- Clasificación
- Asociación
- Agrupamientos

Instalación del sistema (válido para Windows y Linux)

Inicialmente se requieren los siguientes softwares:

- JDK1.4 o superior.
- APACHE TOMCAT 5.5
- Gestor de Bases de Datos MySql

Una vez instalados los softwares anteriores se debe copiar la carpeta de la aplicación para el directorio de publicación del APACHE TOMCAT (normalmente root) y contar con las librerías de JDBC de MySql (incluida en el paquete de instalación) y las librerías de las bases de datos en las cuales tengamos nuestros datos. Nos obstante el sistema permite la conexión por ODBC en el caso de no constar con dichas librerías.

La bases de datos para el mantenimiento del sistema (MySql), se entregan con el paquete, en un fichero texto que contiene las instrucciones SQL, permitiéndose ejecutarlas desde el phpMyAmin o cualquier otro generador de consultas para MySql

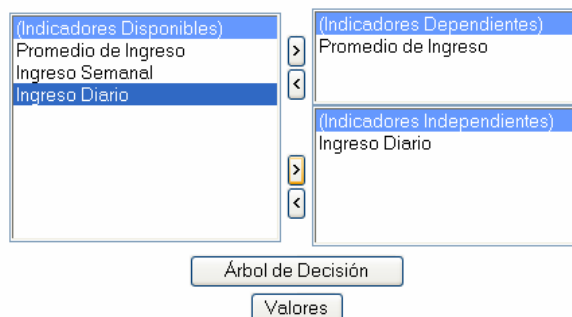
El sistema cuenta con cuatro módulos:

- Minería de Datos
- Consultas a Bases de Datos
- Administración
- Ayuda

### Minería de Datos

Este módulo brinda las posibilidades de utilizar distintos algoritmos de Minería de Datos para el análisis de los datos contenidos en nuestras fuentes de información. Para cada uno de los algoritmos se emplea el mismo punto de partida: una página que permite seleccionar los atributos o parámetros dependientes e independientes que queremos aplicarle el algoritmo.

Figura # 1 Selección de Parámetros



El primer submenú contiene los algoritmos de clasificación, el primero de ellos es la regresión lineal que presenta una opción para generar la ecuación de regresión y otra para pronosticar sobre la base de dicha ecuación. En este algoritmo se debe tener presente que los parámetros suministrados deben ser numéricos

En la opción árboles de decisión crea un gráfico mostrando un árbol de decisión con los datos suministrados, el cual, puede leerse de inicio a fin mediante los colores, el color azul es el inicio del árbol y el rojo la salida o final.

El segundo submenú contiene un algoritmo para crear reglas de asociación (Apriori), que se basa en encontrar en la base de datos suministrada, reglas que tengan una mayor frecuencia, o sea que a partir de los parámetros suministrados se pueda determinar las reglas que más se repiten en el conjunto de datos. En la figura siguiente aparecen dos reglas creadas a partir de datos de prueba suministrados a la aplicación que pueden interpretarse como: si el Costo Total es bajo las Ventas serán Bajas y viceversa. Además se presenta la confiabilidad de cada regla, así como la cantidad de veces que se ha encontrado una regla (número al lado de cada regla)

El tercer submenú se emplea para la separación por grupos o cluster, un método muy eficiente para dividir en grupos afines gran cantidad de información. Por ejemplo se podrían dividir los clientes en diferentes grupos según su frecuencia de compras, etc.

### Consultas a Bases de Datos

En este módulo aparecen dos opciones relacionadas con la generación de consultas o reportes que puedan necesitar los usuarios de la aplicación. En la primera opción se ofrece la posibilidad de solicitar una consulta (información detallada sobre un tema en la base de datos), organizadas por categorías y permitiendo la utilización de parámetros de forma dinámica. En la segunda se permite seleccionar las consultas existentes para ser extraídas de las bases de datos existentes

### Administración

Este módulo está concebido para la administración del sistema, o sea una persona autorizada a cambiar la configuración del mismo. En la primera opción se pueden definir los usuarios que podrán acceder al sistema, así como los accesos a las distintas categorías del mismo.

El próximo submenú (Bases de Datos) se emplea para realizar las conexiones a las bases de datos existentes. De una forma sencilla se permite seleccionar el tipo de conexión a realizar y escoger las tablas o vistas que serán utilizadas por los usuarios de la aplicación. El sistema muestra cada tabla o vista cuando se selecciona y ofrece la posibilidad de dar nombres o alias a los campos y tablas de la base de datos para facilitar la posterior interpretación de éstas por los usuarios. Es importante destacar que la configuración realizada sobre las bases de datos se almacena en la base de datos de la aplicación, logrando así la persistencia de la misma.

## Ayuda

El módulo referente a la ayuda se encuentra en construcción en estos momentos, en él se trata de lograr una referencia interactiva para facilitar el trabajo de los usuarios de la aplicación.

## **Conclusiones**

Como resultado de este trabajo:

- Definimos y estructuramos las informaciones que son necesarias para, de acuerdo a los datos operativos de la entidad, utilizar técnicas de Minería de Datos en el proceso de Toma de Decisiones por parte de los directivos de la corporación Cuba Ron.
- Diseñamos un software, en tecnología WEB y soft libre, que permita la aplicación de técnicas de minado, por parte de los directivos, convirtiéndose en herramienta de trabajo en los procesos de dirección.
- Elaboramos, además, las indicaciones metodológicas que permitan la implementación y aplicación de este trabajo con fines gerenciales.

## **Bibliografía**

- [1] Peña, Pablo Brendell, "To know or not to be". Conocimiento: El oro gris de las organizaciones. Colección Gestión del Conocimiento, Vol. 5, Ed. Fundación DINTEL, 2001.
- [2] Maestre, Pedro Yenes, "Diccionario de Gestión del Conocimiento e Informática". Ed. Fundación DINTEL, 2001.
- [3] La gestión del conocimiento en la nueva economía. Por: Erick Serradell López (eserradell@uoc.edu) y Ángel A. Juan Pérez (ajuanp@uoc.edu). <http://www.uoc.es/>
- [4] Pérez-Bustamante, G.I, "Gestión del Conocimiento e Incidencia Institucional en el proceso de aprendizaje". [www.fcee.ulpgc.es/Acede98/acede/mesa02]
- [5] Ranguelov, S.Y., "La Gestión del conocimiento. Un enfoque para el desarrollo social", 2002. [www.gestiondelconocimiento.com]
- [6] <http://www.cio.com/research/knowledge/edit/kmabcs.html>
- [7] La Gestión del Conocimiento: Nueva Cultura Organizacional, 2001, <http://www.doc6.es>
- [8] Bill Gates, Los negocios en la era digital.
- [9] Varios Autores, "La Gestión del conocimiento: La Herramienta del futuro", Trend Management, No.3, 2000

- [10] "Los sistemas de gestión del conocimiento", 2002, <http://www.canalticom/magazine>
- [11] ¿Qué es un Proyecto de Gestión del Conocimiento?, 2001, <http://www.businessinnovation.ey.com/mko/pdf>
- [12] García, Gerardo Cabrera, "De la era de la información a la sociedad del conocimiento", Revista Ciencia, Innovación y Desarrollo, Vol. 6, No.4, 2001, pág. 17-20.
- [13] Estudio de caso de Gestión del Conocimiento en Ernest Young, 1998, <http://www.mcombs.utexas.edu/kman/>
- [14] Simeón, Rosa Elena, "Cuba posee una nueva riqueza de conocimientos" Revista Nueva Empresa, Vol.1, No.1, 2001, pág.3-5.
- [15] Grupo de Especialistas del CITMA, "Bases para la introducción de la Gestión del conocimiento en Cuba", Revista Ciencia, Innovación y Desarrollo, Vol. 6, No.4, 2001, pág.11-14.
- [16] del Puerto, Roberto. "Estrategia del Ministerio de Informática y las Comunicaciones de Cuba respecto al software Libre", Noviembre 2002. Comisión del software Libre, grupo de Informatización de la sociedad.
- [17] Herrera, Álvaro [alvherre@webhost.cl](mailto:alvherre@webhost.cl). Artículo "La importancia del software libre". Julio de 1999.
- [18] Fuentes Quiroz, Iván. "Aplicaciones de Desarrollo para la Construcción de Sitios Interactivos en Internet para el Comercio Electrónico". Tesis de Licenciatura. Universidad de las Américas-Puebla, Verano 2001.
- [19] Hanna, Phil. "JSP: The Complete Reference". Osborne/McGraw Hill, 2001
- [20] Turegano Molina, Javier. Instalación de un Servidor de Prácticas de Tecnologías de Internet. 01-07-2003. <https://www.linuxalbacete.org/articulos/ta/ta1.htm>
- [21] Diccionario de la Biblioteca de Consulta Microsoft Encarta 2004.
- [22] <http://www.septeto.com/documentos/distribuciones.html>