MÉTODOS VISUALES PARA LA REPRESENTACIÓN DE RESULTADOS DE ALGORITMOS DE MINERÍA DE DATOS.

MSc. Emma Regla Rizo Rizo[1]

Ing. Luis Ernesto Yanes Santiago [2.]

1. Universidad de Matanzas – Sede "Camilo Cienfuegos", Vía Blanca Km.3, Matanzas, Cuba. emma.rizo@umcc.cu

CD de Monografías 2015
(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

Resumen

En el mundo actual la información juega un papel fundamental en el proceso de toma de decisiones por lo que se recopilan grandes cantidades de datos a diario. De estos datos se extrae información utilizando técnicas de minería de datos. Para facilitar el entendimiento de los resultados de minería de datos se utiliza la visualización de datos, de forma tal que se transformen en información comprensible para el usuario y se obtenga el conocimiento necesario para una óptima toma de decisiones. Esta investigación está dirigida a determinar técnicas de visualización que faciliten el entendimiento de los resultados que se obtienen al aplicar algoritmos de minería de datos para las técnicas de clasificación y agrupamiento. Para lograrlo se plantea la necesidad de realizar una búsqueda bibliográfica como soporte teórico de la investigación, estudiar los mecanismos de representación del conocimiento e implementar un *software* que visualice la información. Las pruebas que se realizaron a partir de las técnicas de caja blanca y caja negra, así como las pruebas de aceptación demostraron el buen funcionamiento del *software*. Con el análisis de los casos de estudio se demostró que el sistema posee un correcto funcionamiento y cumple con los requerimientos definidos.

Palabras claves*: Minería de datos, clasificación, agrupamiento, software.*

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

Introducción

En los últimos años se manifiesta un crecimiento considerable de la capacidad del ser humano de generar y colectar datos; el verdadero valor de este cúmulo de datos reside en la información que se le puede extraer, sin embargo las técnicas clásicas de recuperación de la información son insuficientes para esta tarea; por lo que se utiliza la minería de datos que aplica técnicas como la inteligencia artificial para encontrar patrones y relaciones dentro de los datos.

En la actualidad muchos negocios exitosos emplean métodos analíticos avanzados de minería de datos para incrementar ganancias, maximizar la eficiencia operativa, reducir costos y mejorar la satisfacción del cliente. Esto se debe a que la información y el conocimiento que se obtiene de este proceso, además de mejorar la comprensión de los fenómenos que rodean a la sociedad, ayuda en el proceso de toma de decisiones (Vallejos, 2006).

La minería de datos encuentra aplicación en todas las áreas en las que se disponga de información almacenada como son: el área comercial, *marketing*, industria, *internet* y la agricultura. Se aplica a datos numéricos, imágenes de satélite, mamografías, música, archivos de ordenador, entre otras (Sumathi et al., 2006).

La información obtenida al aplicar técnicas de minería de datos no siempre es comprensible para el usuario no experto a quien, por lo general, va dirigido este conocimiento y es por ello que surgen las técnicas de visualización de datos.

La visualización de datos es el proceso de búsqueda, interpretación, contrastación y comparación de datos que permite un conocimiento en profundidad y detalle, de forma tal que se transformen en información comprensible para el usuario y se conviertan en conclusiones y conocimiento para una óptima toma de decisiones (Keim, 2002).

Desarrollo

El Descubrimiento de Conocimiento en Bases de Datos (KDD por sus siglas en inglés) es el proceso que se encarga de extraer la información implícita en los datos mediante la preparación, sondeo, exploración de los mismos y la interpretación de los resultados obtenidos. Permite extraer el conocimiento que da un significado a los patrones encontrados (Vallejos, 2006; Martínez Martínez, 2014).

La Minería de Datos (MD) es frecuentemente tratada como sinónimo del proceso KDD, sin embargo es una etapa del mismo que se encarga de identificar el conocimiento que reside dentro de un conjunto de datos. Para esto utiliza como materia prima las bases de datos y para su procesamiento reúne las ventajas de diferentes áreas como son la Estadística, la Inteligencia Artificial, la Computación Gráfica, las Bases de Datos y el Procesamiento

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

Masivo. Esta disciplina encuentra aplicaciones en detección de intrusiones, análisis del uso de la web, análisis de imagen, detección de comportamiento anormal y *marketing* dirigido.

Técnicas de MD

En la Minería de Datos se utilizan técnicas que dependen del tipo de conocimiento que se desea extraer. Una forma de clasificar estos tipos de conocimiento y técnicas se puede observar la Tabla 1. (Miller et al., 2009)

Tipo de Conocimiento	Descripción	Técnicas
Segmentación o agrupamiento	Determinar un conjunto finito de grupos implícitos que describe los datos.	- Análisis de grupo (agrupamiento)
Clasificación	Predecir una clase perteneciente a un conjunto de datos basado en una fuente de datos de entrenamiento.	- Clasificación Bayesiana - Árbol de decisión - Redes neuronales
Asociación	Buscar relaciones entre reglas de asociación y correlación, o predecir el valor de un atributo basado en el valor de otros atributos.	- Reglas de asociación - Redes Bayesianas
Desviación	Buscar elementos de los datos que muestren una inusual desviación de los esperados.	- Agrupamiento y otros métodos de MD - Detección de la parte aislada - Análisis de evolución
Tendencias y análisis de regresión	Líneas y curvas resumiendo la base de datos, a menudo, todo el tiempo.	- Regresión - Extracción de patrones secuenciales
Generalización	Descripción concisa de los	- Reglas de síntesis

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

	datos	- Inducción (aislamiento) orientada a atributo

Tabla1. Técnicas de minería de datos.

El desarrollo de esta investigación se centra en dos de las técnicas de minería de datos más importantes: agrupamiento y clasificación. Los patrones o conocimientos que se extraen con estas dos técnicas brindan comúnmente el tipo de información más solicitada por los usuarios. El resto de las técnicas se quedan como objetivo de futuros trabajos. Visualización de datos

La visualización de datos se encarga de representar gráficamente los datos con el propósito de ayudar al usuario a entender la información que reside en dichos datos.

Para varios autores la visualización de datos se puede definir como un proceso de representación gráfica de los datos y la interacción con dicha representación para obtener una visión de la información contenida en los datos. Tiene como objetivo principal comunicar información clara y efectiva a través de recursos gráficos (Friendly, 2008; Simoff *et al*., 2008).

El proceso KDD presenta muchas dificultades para su automatización debido a que necesita de inteligencia de alto nivel (es decir inteligencia humana). En este punto la visualización se convierte en una herramienta poderosa para integrar la inteligencia y el conocimiento humano dentro del proceso KDD. La alta efectividad del sistema visual humano en el reconocimiento de patrones, tendencias y anomalías, se puede aprovechar en varias etapas del proceso KDD como son: la formulación de consultas, la selección de técnicas y la interpretación de los resultados (Miller et al., 2009).

La exploración visual de datos generalmente sigue tres pasos: primera revisión, enfoque y filtrado, y detallar según se necesite. Primero el usuario identifica patrones de interés o grupos en los datos y se centra en uno o varios de ellos. Para analizar los patrones el usuario debe sondear y acceder a los detalles de los datos. En el segundo paso es importante mantener la visión general de los datos, mientras el usuario se enfoca en el subconjunto utilizando otra técnica de visualización. En el tercer paso para una exploración más detallada de los subconjuntos de interés, el usuario necesita una capacidad de exploración que corresponda al nivel de detalle que se quiere alcanzar (Keim et al., 2002; Stahl *et al*., 2013).

La visualización de datos se basa en la aplicación del algoritmo de minería de datos y en la técnica visual. Existen tres corrientes que integran al humano en el proceso de exploración de los datos Anexo 2 (Keim *et al*., 2004).

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

1. Visualización Precedente (*Preceding Visualization* (PV)): Los datos se visualizan antes de aplicar el algoritmo de minería de datos. El analista tiene total control sobre el análisis en el espacio de investigación debido a la interacción con los datos antes de ser procesados. Se descubren patrones interesantes a partir de la exploración de los datos.
2. Visualización Subsiguiente (*Subsequent Visualization* (SV)): Se ejecuta el algoritmo de minería de datos para extraer patrones de la fuente de datos. Se visualizan esos patrones de modo que sean interpretables por el analista, que se basa en esta representación visual, para decidir si es necesario volver a aplicar el algoritmo de minería de datos y utilizar parámetros de entrada diferentes para obtener mejores resultados.
3. Visualización Estrechamente Integrada (*Tightly Integrated Visualization* (TIV)): Se ejecuta el algoritmo de minería de datos pero no obtiene el resultado final. Se utiliza una técnica de visualización para representar el resultado intermedio del proceso de exploración de datos. Esto le permite al usuario especificar la próxima tarea de minería de datos a ejecutar. Luego el analista identifica patrones interesantes a partir de la representación visual del resultado intermedio basado en sus conocimientos.

Los gráficos son útiles en la representación de relaciones entre conjuntos de datos. Por ejemplo los dendogramas se utilizan para algoritmos de agrupamiento jerárquico, para los problemas de clasificación y regresión se usan árboles y los diagramas de ruta para redes Bayesianas o para expresar dependencias entre diferentes variables (Chen *et al*., 2008).

Tecnologías utilizadas.

Para el desarrollo de esta aplicación *web*, se utilizaron las siguientes herramientas, tecnologías y metodologías de desarrollo existentes.

La metodología de desarrollo Programación Extrema (XP por su siglas en inglés), se centra en potenciar las relaciones interpersonales. Se basa en la retroalimentación continua entre el cliente y el equipo de desarrollo, comunicación fluida entre todos los participantes, simplicidad en las soluciones implementadas y coraje para enfrentar los cambios (Campos et al., 2015).

Weka

Es un conjunto de bibliotecas escritas en *Java* para la extracción de conocimientos de bases de datos. Es un software que se desarrolla bajo la licencia publica Genera (GPL por sus siglas en inglés). Posee distintas herramientas de minería de datos como reglas de asociación, agrupamiento, clasificación y regresión. (Bouckaert *et al*., 2014), se considera software libre porque es un programa licenciado en términos que garantizan a sus usuarios el derecho de ejecutarlos, copiarlos, distribuirlos, estudiarlos, cambiarlos y mejorarlos (Zanotti, 2016).

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

Arquitectura Cliente/Servidor.

La arquitectura Cliente/Servidor se divide en dos partes claramente diferenciadas. El cliente produce una demanda de información o solicita recursos y el servidor recibe la demanda y responde en consecuencia. (Gironés, 2012).

Lenguajes de Programación para la Web.

El presente software se desarrolló utilizando generador de proyectos JHipster, que crea una aplicación web utilizando patrón de arquitectura MVC (Modelo Vista Controlador) y buenas prácticas como el desacoplamiento de código. Se utilizó Java (Stärk *et al*., 2012), uno de los lenguajes de programación más utilizados en la actualidad, para desarrollar la aplicación del lado del servidor y para el cliente se utilizó AngularJS, el marco de trabajo JavaScript más popular en la actualidad. (Jain *et al.*, 2015)

MySQL

Es un gestor de base de datos sencillo, rápido, sólido y flexible. Es idóneo para la creación de bases de datos con acceso desde páginas *web* dinámicas. Es gratis para aplicaciones no comerciales y soporta gran cantidad de tipos de datos (Schwartz *et al.*, 2012).

Servidor Web(Apache Tomcat)

Es un *software* desarrollado con *Java*, por lo que puede funcionar en cualquier sistema operativo con una máquina virtual *Java* correspondiente, que sirve como servidor *web* con soporte de *servlets* y *JSPs*. *Apache Tomcat* se desarrolla en un entorno abierto y de participación, bajo la licencia de *Apache Software License*. (Pérez Díaz, 2016).

Solución propuesta.

En este trabajo se desarrollará una aplicación para determinar si los métodos visuales que se proponen, permiten una mejor comprensión de los resultados obtenidos. Las técnicas de interés en este trabajo son clasificación con el algoritmo C4.5 (o J48 por su nombre en WEKA) y agrupamiento con el algoritmo *k-means* (o *SimpleKMeans* por su nombre en WEKA).

Se escogió el algoritmo C4.5 porque puede manejar instancias ponderadas por pesos, procesar datos categóricos, permite utilizar árbol podado o no podado, optimiza el rendimiento en un conjunto fijo y además permite la construcción de árboles binarios. El algoritmo *k-means* se seleccionó porque además de ser uno de los algoritmos de agrupamiento con más popularidad, flexibilidad y aplicabilidad, presenta características como: minimizar la dispersión de los grupos gracias a la reasignación iterativa de los

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

miembros de un grupo y su complejidad tiempo y espacio son pequeñas en relación con otros algoritmos de agrupamiento.

El sistema propuesto utilizará fuentes de datos en formato arff. Para futuros trabajos se puede implementar funcionalidades que permitan operar con datos provenientes de bases de datos u otros tipos de ficheros, como rdf o xml.

El sistema permitirá copiar la fuente de datos en el servidor y partir de este punto, la aplicación se encarga de extraer los metadatos necesarios de la fuente de datos y almacenarlos en base de datos.

Primero el usuario escogerá la fuente de datos a la que desee extraerle el conocimiento y comenzará a responder a una serie de preguntas, que le indicará al sistema la técnica que se necesita aplicar. La aplicación ejecutará el algoritmo correspondiente y almacenará los resultados.

A continuación el sistema procesa los resultados para representarlos a partir del método visual según el algoritmo aplicado. Por último el *software* grafica los resultados y lo muestra al usuario para que analice la información de manera visual.

Métodos visuales

Sectores en Sectores

El método visual Sectores en Sectores (SonS por sus siglas en inglés) propuesto en Martínez, 2014, se basa en extraer información visual de grupos de datos, representando la cantidad de instancias en cada grupo, el valor de los centroides de dichos grupos y las relaciones que existen entre varios grupos y variables. Utiliza para su representación visual un gráfico de pastel.

Este método consta de tres pasos

1. División de una circunferencia en tantos sectores como grupos existentes: Cada segmento de pastel corresponde a cada grupo resultante de la aplicación del algoritmo. El arco de cada sector es proporcional al número de patrones que contienen los mismos. De esta forma se reconoce fácilmente la relevancia de cada grupo (Figura 2.4).
2. División de las porciones del pastel en dependencia de la cantidad de atributos y los valores de cada uno: La parte interna corresponde a la primera variable y así sucesivamente. El radio de cada parte (r_i) depende del valor relativo de cada variable con respecto a la suma de todos ellos, como se puede observar en la ecuación 1.

$$r_i = \frac{|x_i|}{\sum_{k=1}^{N} |x_k|}, i = 1 \dots N \qquad \text{Ec.1}$$

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

Donde x_i es el centroide de cada variable de un grupo. Ver Anexo 4.

3. Asignación de colores para identificar el valor real de los centroides: Se agrega una barra de colores con una columna por cada variable. Al centroide de cada variable se le asigna un color.

Al utilizar este método se consideraron algunas modificaciones para hacer más exacta la representación de los datos. En lugar de utilizar la relación entre los colores de cada subsector con la barra de colores para especificar los valores de los datos; el usuario podrá ver los valores correspondientes a cada variable señalando el subsector adecuado. De esta forma puede identificar más rápidamente qué valor tiene la variable en un determinado grupo.

El método visual que se propone para visualizar los resultados que brinda el algoritmo C4.5 es la representación de un árbol de decisión; donde se utiliza una elipse para mostrar los nombres de las variables de la fuente de datos, una línea para representar los posibles valores de cada variable y un rectángulo para identificar las hojas o instancias clasificadas.

Validación de la solución propuesta

Las pruebas al *software* que permiten conocer el grado de calidad del producto, y de esta forma, comprobar el cumplimiento de los requisitos funcionales del sistema desarrollado. Se prueban todas las características de la aplicación con el fin de verificar la fiabilidad y calidad de la aplicación como un todo.

En la metodología XP la fase de pruebas cumple un papel importante, ya que se convierten en una herramienta de desarrollo y no un paso de verificación que puede despreciarse.

Los principios básicos que guían el buen funcionamiento de las pruebas de *software* y que es necesario conocerlos antes de aplicar algún método de prueba son (Pressman, 2011):

- Que a todas las pruebas se les debería poder hacer un seguimiento hasta los requisitos del cliente.
- Deben planificarse mucho antes de que empiecen.
- Deben comenzar por lo pequeño y progresar hacia lo grande.

- Se realizaron pruebas de caja blanca, pruebas de caja negra y pruebas de aceptación para garantizar la calidad del software Ver Anexo 1.

Casos de estudios.

En este epígrafe se explicarán dos casos de estudio aplicando la técnica de clasificación y de agrupamiento respectivamente.

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

Caso de estudio aplicando la técnica de clasificación

El primer conjunto de datos a estudiar se trata de un conjunto de datos nominales. En concreto es una base de datos de un zoológico imaginario creado por Richard Forsyth en 1990. Distingue 7 clases de animales, 17 atributos (y además el nombre de la especie, único para cada muestra), 15 booleanos y 2 numéricos, y no faltan valores de ningún atributo. Los atributos booleanos pueden tomar dos valores, verdadero o falso.

Tiene una amplia gama de animales clasificados en mamíferos, aves, reptiles, peces, anfibios, insectos y crustáceos; y consta de 101 instancias (que se corresponden con 101 tipos de animales).

A continuación se muestra el resultado que se obtiene de aplicar el algoritmo C4.5 utilizando la biblioteca Weka (Figura 1):

```
feathers = false
|   milk = false
|   |   backbone = false
|   |   |   airborne = false
|   |   |   |   predator = false
|   |   |   |   |   legs <= 2: invertebrate (2.0)
|   |   |   |   |   legs > 2: insect (2.0)
|   |   |   |   predator = true: invertebrate (8.0)
|   |   |   airborne = true: insect (6.0)
|   |   backbone = true
|   |   |   fins = false
|   |   |   |   tail = false: amphibian (3.0)
|   |   |   |   tail = true: reptile (6.0/1.0)
|   |   |   fins = true: fish (13.0)
|   milk = true: mammal (41.0)
feathers = true: bird (20.0)
```

Figura 1 Resultado del algoritmo C4.5.

Con la aplicación del *software* se obtiene una representación visual del árbol de decisión que se puede observar en la figura 2

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

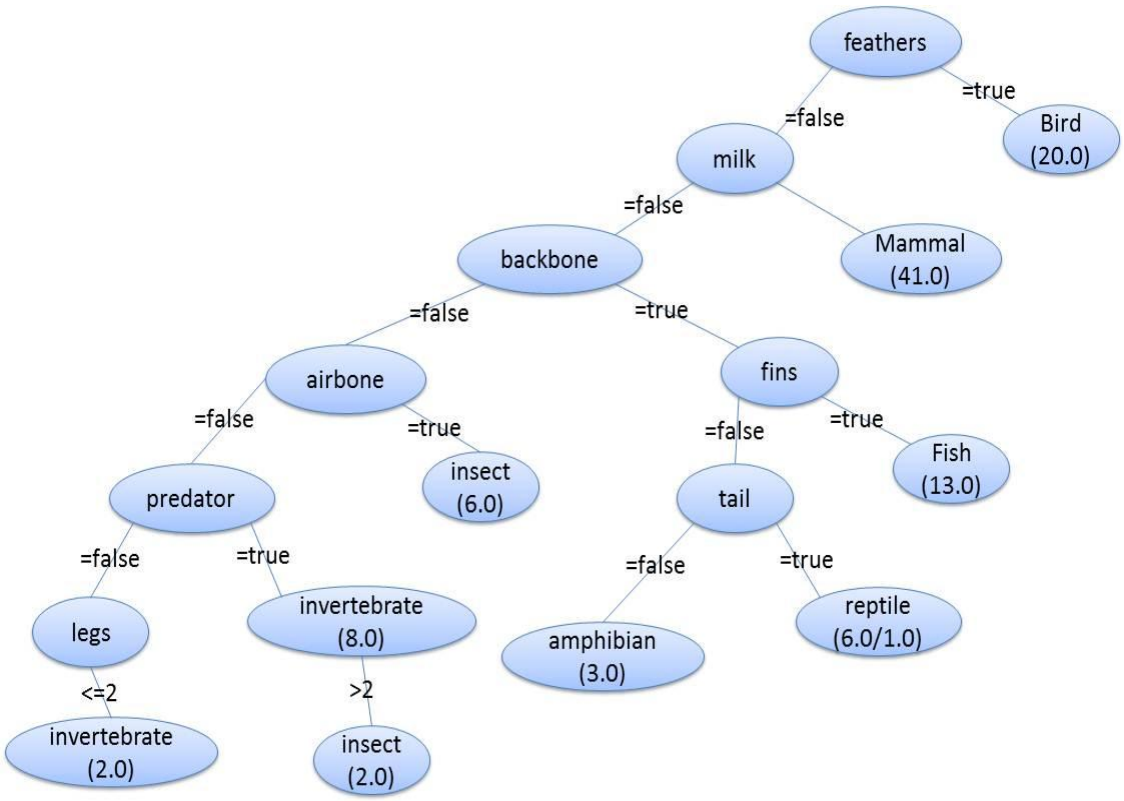

Figura 2. Visualización del árbol obtenido con el algoritmo C4.5.

Caso de estudio aplicando la técnica de agrupamiento

El segundo conjunto de datos a estudiar está compuesto por datos numéricos. Corresponden a investigación que se realiza en la Empresa de Proyecto de Arquitectura e ingeniería de Matanzas (EMPAI) titulado "Modelo de predicción para la planificación objetiva de los insumos en la EMPAI.", para la visualización de los datos.

A continuación se expone como la biblioteca Weka brinda el resultado al aplicar la técnica de clustering en la figura 3:

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

```
Number of iterations: 10
Within cluster sum of squared errors: 19.053889098350616
Missing values globally replaced with mean/mode
Cluster centroids:
Cluster#
Attribute Full Data 0 1 2 3 4 5
(88) (6) (5) (22) (22) (27) (6)
==========================================================
mes1 50.0114 33.8333 49.4 44.3409 51.25 56.5556 53.5
mes2 49.7386 30.6667 51.4 49.9091 47.0909 55.8519 49
mes3 49.7045 33.6667 49.2 47.4773 49.4318 55.9259 47.3333
mes4 49.6364 41.5 43.4 44.1818 51.1818 55.5926 50.5
mes5 49.3977 42.1667 36.2 45.7727 50.1364 55.1481 52.3333
mes6 49.3409 46.5 32.4 48.5682 47.6136 54.7778 51
mes7 49.3523 50 28.8 47.0455 49.8636 54.6667 48.5
mes8 49.2955 49 34.6 49.5682 47.75 55 40.8333
mes9 49.1591 48.75 45.6 44.5682 51.6364 55.6667 31
actual 48.9432 48.3333 44 45.0455 50.7727 55.5185 31.6667
```

Figura 3. Resultado técnica de clustering.

Al utilizar el método SonS los resultados quedan representados como se muestra en la figura 4

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

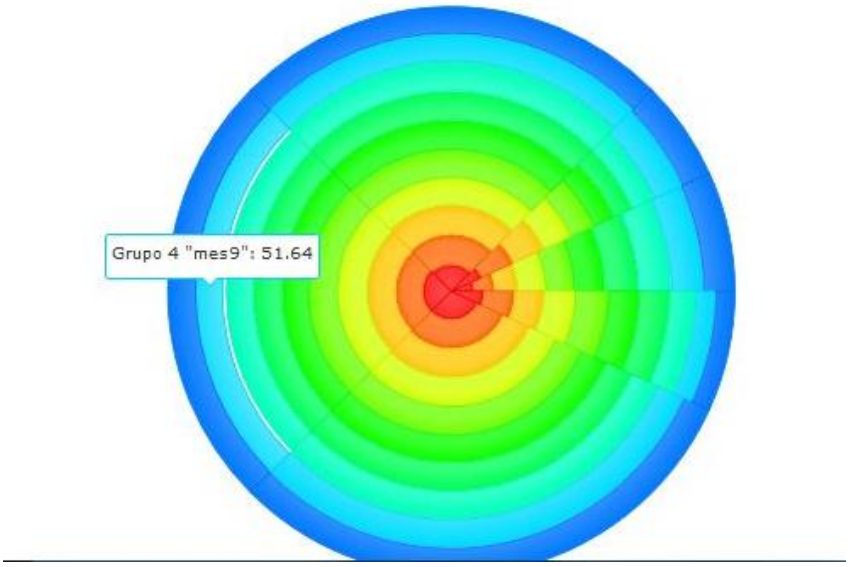

Figura 4 Método visual SonS

Análisis de los casos de estudio

Al establecer una comparación entre los resultados que se obtienen de aplicar algoritmos de Minería de Datos y los resultados que brinda el *software*, se puede apreciar que la representación gráfica de dichos resultados ayuda al usuario a comprender la información que reside en los datos; ya que le permite lidiar fácilmente con datos no-homogéneos, no requiere el entendimiento de complejos algoritmos matemáticos o estadísticos y provee una visión cualitativa de los datos, permitiendo aislar datos anómalos para un análisis más detallado.

Conclusiones

Con el desarrollo de esta investigación se arribó a las siguientes conclusiones:

La determinación de técnicas de visualización de los resultados que se obtienen al aplicar algoritmos de clasificación y agrupamiento, facilita una toma de decisiones mejor fundamentada para el usuario inexperto, lo cual corrobora la hipótesis planteada.

Las técnicas de clasificación y agrupamiento brindan comúnmente el tipo de información más solicitada por los usuarios.

El desarrollo de la aplicación *web* utilizando el lenguaje de programación *Java* y el generador de proyectos *JHipster*, hizo posible aplicar los métodos visuales: Árbol de decisión y SonS modificado.

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X

Con el análisis de los casos de estudio se demostró que el sistema posee un correcto funcionamiento y cumple con los requerimientos definidos en un principio, para ayudar al usuario a entender la información que se extrae de una fuente de datos al aplicar las técnicas de minería de datos de agrupamiento y clasificación.

Bibliografía (Ejemplo Norma ISO)

VALLEJOS, S.J. Minería de Datos [Trabajo de Adscripción]. [Corrientes-Argentina]: Universidad Nacional del Nordeste; 2006. 31 p.

SUMATHI, S.; SIVANANDAM, S.N. Introduction to Data Mining and its Applications. Vol. 29; 2006. 800 p.

KEIM, D.A. Information Visualization and Visual Data Mining. In: GRAPHICS, I.T.O.V.A.C., editor. Vol. 7; 2002.

MARTÍNEZ MARTÍNEZ, J.M. Visual Data Mining: Real Applications and New Approaches [Doctoral]. [Valencia]: University of Valencia; 2014. 263 p.

MILLER, H.J.; HAN, J. Geographic Data Mining and Knowledge Discovery: CRC Press; 2009.

FRIENDLY, M. Milestones in the history of thematic cartography, statistical graphics, and data visualization. Seeing Science : Today American Association for the dvancement of Science. 2008.

SIMOFF, S.; B¨OHLEN, M.; MAZEIKA, A. Visual Data Mining: Theory, Techniques and Tools for Visual Analytics.: Springer.; 2008.

KEIM, D.; WARD, M. Visual Data Mining Techniques. University of Konstanz, Germany and Worcester Polytechnic Institute, USA. 2002.

STAHL, F.; GABRYS, B.; MEDHAT GABER, M.; BERENDSEN, M. An Overview on Interactive Visual Data Mining Techniques for Knowledge Discovery. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2013 22, February, 2016.

KEIM, D.; SIPS, M.; ANKERST, M. The Visualization Handbook. USA: The Boeing Company; 2004. 813 p.

CHEN, C.-H.; HARDLE, W.; UNWIN, A. Handbook of Data Visualization. USA: Springer-Verlag TELOS; 2008.

CAMPOS, S.G.; MARTÍNEZ, L.F.F. Programación Extrema: Prácticas, Aceptación y Controversia. CULCyT. 2015 (15).

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"

ISBN: XXX-XXX-XX-XXXX-X

ZANOTTI, A. El software libre y su difusión en Argentina: mercado, Estado, sociedad. POLIANTEA. 2016;11(21):147-166.

BOUCKAERT, R.R.; FRANK, E.; HALL, M.; KIRKBY, R.; REUTEMANN, P.; SEEWALD, A.; SCUSE, D. WEKA Manual for Version 3-7-12. New Zealand: University of Waikato; 2014.

GIRONÉS, J.T. El gran libro de Android: Marcombo; 2012.

STÄRK, R.F.; SCHMID, J.; BÖRGER, E. Java and the Java virtual machine: definition, verification, validation: Springer Science & Business Media; 2012.

SCHWARTZ, B.; ZAITSEV, P.; TKACHENKO, V. High performance MySQL: Optimization, backups, and replication: " O'Reilly Media, Inc."; 2012.

PÉREZ DÍAZ, A.J. Tomcat, Apache Tomcat, Jakarta Tomcat; [cited 2016 25, abril]. Available from: http://www.ajpdsoft.com/modules.php?name=Encyclopedia&op=content&tid=769

PRESSMAN, R.S. Técnicas de prueba del software, en Ingeniería del Software. Un enfoque práctico.; 2011.

(c) 2015, Universidad de Matanzas "Camilo Cienfuegos"
ISBN: XXX-XXX-XX-XXXX-X