

# SLD070 COMPONENTE WEB PARA EL ANÁLISIS DE INFORMACIÓN CLÍNICA USANDO LA TÉCNICA DE MINERÍA DE DATOS AGRUPAMIENTO

## SLD070 WEB COMPONENT FOR THE ANALYSIS OF CLINICAL INFORMATION USING THE TECHNIQUE OF CLUSTERING DATA MINING

Ing. Alexeis Joel Ochoa Reyes<sup>1</sup>, Ing. Arturo Orellana García<sup>2</sup>, Ing. Yovannys Sánchez Corales<sup>3</sup>, Ing. Frank Davila Hernández<sup>4</sup>

1 Centro de Informática Médica, Universidad de las Ciencias Informáticas, Cuba, [ajchoa@uci.cu](mailto:ajchoa@uci.cu), Carretera San Antonio de los Baños Km 2 ½, Boyeros, La Habana, Cuba.

2 Centro de Informática Médica, Universidad de las Ciencias Informáticas, Cuba, [aorellana@uci.cu](mailto:aorellana@uci.cu).

3 Centro de Informática Médica, Universidad de las Ciencias Informáticas, Cuba, [yscorales@uci.cu](mailto:yscorales@uci.cu).

4 Centro de Informática Médica, Universidad de las Ciencias Informáticas, Cuba, [fdavila@uci.cu](mailto:fdavila@uci.cu).

**RESUMEN:** La digitalización de los diferentes procesos y la automatización de los servicios generan grandes volúmenes de información. La Minería de Datos (MD) es una técnica de Inteligencia Artificial que permite encontrar la información no trivial que residen en los datos almacenados. La presente investigación pretende desarrollar una vista de análisis para el Sistema Integral para la Atención Primaria de Salud (SIAPS), usando la técnica de agrupamiento enmarcada en el algoritmo Simple K-Means, con el objetivo de realizar un análisis de la información clínica de los pacientes; para ello se plantea la extracción del conocimiento del almacén de datos alimentado del repositorio de historias clínicas electrónicas. La investigación se sustenta en la herramienta de libre distribución WEKA, esta funciona de forma aislada al SIAPS; la interfaz, así como las vistas, modelos e informes generados por WEKA en ocasiones resultan de difícil comprensión por los profesionales de la salud, los que no necesariamente tienen que poseer conocimientos avanzados de las nuevas tecnologías de la información. Para el desarrollo de la solución se empleó el lenguaje de programación Java 1.6, como servidor de aplicación JBoss 4.2 y Eclipse 3.4 como plataforma de desarrollo, como Sistema Gestor de Bases de Datos PostgreSQL 8.4 y SEAM como framework de integración. Durante todo el proceso se hizo uso de la plataforma Java Enterprise Edition 5.0. Como resultado se espera obtener una vista de análisis que facilite la comprensión de los modelos generados, apoyando de esta forma el proceso de toma de decisiones clínicas.

**Palabras clave:** Almacén de datos, extracción del conocimiento, Inteligencia Artificial, Minería de Datos, Simple K-Means, vista de análisis.

**ABSTRACT:** The digitization of the different processes and automation services generate large volumes of information. Data mining (DM) is an artificial intelligence technique that allows finding non-trivial information residing in stored data. This research aims to develop a view of analysis for the Integral System for Primary Health Care (SIAPS), using grouping technique framed on Simple K-Means algorithm, with the goal of completing an analysis of the patients' clinical information, for it raises the extraction of knowledge from data warehouse powered by the repository of electronic medical records. The research is based on the free distribution tool WEKA, it works in isolation of SIAPS, the interface, as well as the views, models and reports generated by WEKA are sometimes difficult to understand by health professionals, who do not necessarily have to possess advanced knowledge of new information technologies. For the development of the solution was used Java 1.6 as a programming language, JBoss 4.2 as the application Server and Eclipse 3.4 as a development platform. Java Enterprise Edition 5.0 platform was used during the whole process. An analysis view to facilitate the understanding of the generated models is expected as a result, to support the process of making clinical decisions.

**KeyWords:** Data warehouse, knowledge extraction, Artificial Intelligence, Data Mining, Simple K-Means, analysis view.

## 1. INTRODUCCIÓN

Con el desarrollo de la Inteligencia Artificial (IA) y las técnicas de Minería de Datos (MD) pueden ser procesados grandes volúmenes de información con el objetivo de extraer patrones que residan en los datos almacenados. La MD es una tecnología novedosa que integra diferentes técnicas de análisis de datos y extracción de modelos. La posibilidad de extraer patrones, describir tendencias y regularidades, predecir comportamientos y en general, aprovechar la utilidad de la información almacenada, comúnmente heterogénea y en grandes cantidades, permite a individuos y organizaciones analizar, entender y modelar de una manera más eficiente y precisa el contexto en el que deben actuar y tomar decisiones [1].

Dentro de las técnicas de MD existentes se encuentra el agrupamiento (*Clustering*). Los algoritmos utilizados por la técnica de agrupamiento permiten clasificar un conjunto de elementos de muestra en un determinado número de grupos, basándose en las semejanzas y diferencias existentes entre los componentes de la muestra [2], o sea, consiste en obtener grupos naturales a partir de los datos, de forma tal que los objetos de un mismo grupo son muy similares entre sí, y al mismo tiempo, son muy diferentes a los objetos de otros grupos [3].

Disminuir el error médico, mejorar los procesos de salud y garantizar el cuidado de los pacientes ha sido foco de preocupación constante de todos los miembros del equipo de salud. En este contexto surgen los Sistemas Clínicos de Soporte para la Toma de Decisiones (*Clinical Decision Support System* – por sus siglas en inglés CDSS) los que constituyen componentes fundamentales del proceso que conlleva la informatización de la capa clínica. Estos sistemas se basan en proveer al médico o cualquier individuo de información específica del paciente o población, inteligentemente procesada en el momento preciso para garantizar una mejor atención y optimización de los procesos de salud pública [4].

Las Historias Clínicas Electrónicas (HCE) pertenecientes al baluarte del SIAPS se encuentran almacenadas en un repositorio, su información se envía periódicamente a un almacén de datos; lugar donde por el gran volumen de información se hace engorrosa la identificación de patrones comunes, asociaciones y reglas generales de comportamiento que faciliten a los usuarios la toma de decisiones.

Uno de los componentes fundamentales del SIAPS es el módulo para la Toma de Decisiones Clínicas, algunas de sus potencialidades están limitadas por la incapacidad de no contar con mecanismos incorporados que permitan aplicar técnicas de MD a la información clínica almacenada de los pacientes,

con el objetivo que los usuarios encargados de las tareas de toma de decisiones clínicas, puedan visualizar y comprender correctamente los modelos e informes arrojados por las técnicas aplicadas y así facilitar el proceso de toma de decisiones clínicas.

Para el análisis de la información y la clasificación de los pacientes en cuanto a las similitudes de los factores de riesgos a partir de la información clínica almacenada, es usada una herramienta de software libre llamada WEKA. Esta herramienta funciona de forma aislada al SIAPS, trayendo consigo que los usuarios posean dos herramientas de trabajo con entorno y características completamente diferentes, dificultando así el desempeño de estos. Independientemente de esta situación, la interfaz, así como las vistas, modelos e informes generados por WEKA en ocasiones resultan de difícil comprensión por los profesionales de la salud, los que no necesariamente tienen que poseer conocimientos avanzados de las nuevas tecnologías de la información.

En función de facilitar la comprensión de los modelos generados mediante la técnica de agrupamiento por los usuarios que intervienen en el proceso de toma de decisiones clínicas se propone la siguiente interrogante ¿Cómo integrar al Sistema Integral para la Atención Primaria de Salud las diferentes funcionalidades contenidas en la técnica de agrupamiento e implementadas por la herramienta WEKA?

Para dar respuesta a la situación anterior, la presente investigación toma como objeto de estudio, las técnicas de Minería de Datos implementadas por la herramienta WEKA, siendo el campo de acción la técnica de agrupamiento. Se concibió como objetivo general desarrollar una vista de análisis para el Sistema Integral para la Atención Primaria de Salud, usando la técnica de agrupamiento enmarcada en el algoritmo Simple K-Means.

Como resultado de la investigación se pretende desarrollar un componente web basado en la técnica de agrupamiento, el mismo tendrá un aporte práctico sobre el SIAPS, permitiendo que los especialistas encargados de las tareas de toma de decisiones clínicas puedan visualizar y comprender los modelos e informes generados mediante dicho componente, convirtiendo al SIAPS en un sistema más robusto que permitirá agilizar el proceso de análisis y comprensión de la información almacenada.

## 2. CONTENIDO

Existen términos que se utilizan frecuentemente como sinónimos de la minería de datos. Uno de ellos se conoce como "análisis (inteligente) de datos" [5] que suele hacer un mayor hincapié en las

técnicas de análisis estadístico. Otro término muy utilizado, y el más relacionado con la minería de datos, es la extracción o "descubrimiento de conocimiento en bases de datos" o *Knowledge Discovery in Databases* o KDD, según sus siglas en inglés [6].

Aunque algunos autores usan los términos Minería de Datos y KDD indistintamente, como sinónimos, existen claras diferencias entre los dos. Así la mayoría de los autores coinciden en referirse al KDD como un proceso que consta de un conjunto de fases, una de las cuales es la minería de datos [7]. De acuerdo con esto, el proceso de minería de datos consiste únicamente en la aplicación de un algoritmo para extraer patrones de datos y se llamará KDD al proceso completo que incluye pre-procesamiento, minería y post-procesamiento de los datos.



Figura 1. Procesos que involucra el Descubrimiento de Conocimiento [8].

El KDD según Fayyad [9] es la extracción automatizada de conocimiento o patrones interesantes, no triviales, implícitos, previamente desconocidos, potencialmente útiles y predictivos de la información de grandes bases de datos.

La Minería de Datos (MD) o por su nombre en inglés *Data Mining* es el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos [10]. Las herramientas de Minería de Datos predicen futuras tendencias y comportamientos, permitiendo en los negocios la toma de decisiones.

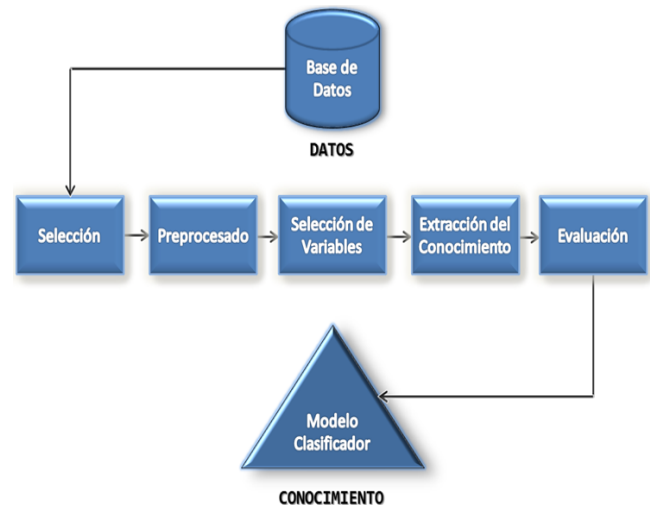


Figura 2. Fases de un proceso de MD [11].

## 2.1 Materiales y Métodos

### 2.2.1 Herramientas y tecnologías utilizadas

#### Eclipse Ganymede 3.4.2

Eclipse es un IDE (del inglés, *Integrated Development Environment*) de código abierto y multiplataforma que ha alcanzado un alto grado de madurez en el desarrollo de lo que se conoce como "Aplicaciones de cliente enriquecido". Cuenta con herramientas para desarrollar aplicaciones de consola, web y servicios web con diferentes servidores de aplicaciones tales como JBoss, Websphere y Glassfish. Fue desarrollado originalmente por IBM (del inglés, *International Business Machines*) y su futuro está ahora en manos de la Fundación Eclipse, una organización independiente sin ánimo de lucro que fomenta una comunidad de código abierto y un conjunto de productos complementarios, capacidades y servicios.

En cuanto a las aplicaciones clientes, Eclipse provee al programador con frameworks muy ricos para el desarrollo de aplicaciones gráficas, para la definición y manipulación de modelos de software, aplicaciones web, entre otros.

Eclipse Ganymede 3.4.2 asegura robustez y rendimiento. Mejora, respecto a versiones anteriores, su soporte de búsqueda con expresiones regulares. Cuenta con un visor de problemas en el código, con asistente para la conversión a *StringBuffer* y con mejoras en el *debugger*.

#### PgAdmin III 1.10.5

PgAdmin III es una herramienta para la administración gráfica de PostgreSQL. Funciona sobre casi todas las plataformas. Fue diseñado para responder a las necesidades de todos los usuarios, desde la escritura de simples consultas SQL a la elaboración

de bases de datos complejas. La interfaz gráfica es compatible con todas las características de PostgreSQL y facilita la administración. La aplicación también incluye un editor de la sintaxis SQL, un editor de código del lado del servidor, un agente para la programación de tareas «SQL/batch/shell» y soporte para el motor de replicación Slony-I. PgAdmin III 1.10.5 es la herramienta seleccionada para la administración gráfica de PostgreSQL 8.4 en el desarrollo de la presente investigación.

### Visual Paradigm para UML

Visual Paradigm para UML (del inglés, *Unified Modeling Language*) es una herramienta CASE (del inglés, *Computer Aided Software Engineering*) aplicable en todo el ciclo de vida del desarrollo de software. Soporta UML, SysML (del inglés, *Systems Modeling Language*), BPMN (del inglés, *Business Process Modeling Notation*), entre otras tecnologías. Permite dibujar todos los tipos de diagramas de clases, generar código desde diagramas y generar documentación. También proporciona abundantes tutoriales UML, demostraciones interactivas de UML y proyectos UML. Presenta licencia gratuita y comercial. Es fácil de instalar y actualizar y compatible entre ediciones.

Visual Paradigm 6.4 ha sido la herramienta seleccionada para soportar el ciclo de desarrollo del componente web para el módulo de Toma de Decisiones del SIAPS. Los cuales lo han seleccionado por su reputación, experiencia, facilidad de los productos y servicios brindados.

### Weka V3.6.2

WEKA (del inglés, *Waikato Environment for Knowledge Analysis*) es una herramienta visual de libre distribución (licencia GNU) desarrollada por un equipo de investigadores de la Universidad de Waikato, Nueva Zelanda. Como entorno de minería de datos conviene destacar:

**Acceso a datos:** Los datos son cargados desde un archivo en formato ARFF (archivo plano organizado en filas y columnas). El usuario puede observar en los diferentes componentes gráficos, información de interés sobre el conjunto de muestras (talla del conjunto, número de atributos, tipo de datos, medidas y varianzas de los atributos numéricos, distribución de frecuencias en los atributos nominales, etc.) (Corría y Shelton, 2004).

**Visualización:** La interfaz gráfica se compone de 4 entornos: Explorer, Consola (CLI), Experimenter y Knowledge Flow.

### Servidor de aplicaciones

JBoss es un servidor de aplicaciones Java EE (del inglés, *Java Enterprise Edition*) de software libre implementado en Java puro. Al estar basado en Java, puede ser utilizado en cualquier sistema operativo que lo soporte. JBoss *Application Server* 4.2.2 ha sido seleccionado para el desarrollo del compo-

nente web que se pretende desarrollar con la presente investigación. El mismo proporciona una gama completa de prestaciones para Java EE 5, así como ampliación de los servicios empresariales, incluyendo clustering, caching y persistencia. JBoss es ideal para aplicaciones Java y aplicaciones basadas en la web. También soporta EJB 3.0 del inglés, (*Enterprise Java*) y esto hace que el desarrollo de las aplicaciones sea mucho más simple.

Una de las facilidades que este servidor presenta es que puede ser instalado sobre varios sistemas operativos, tales como Windows o GNU/Linux.

### eXtensible HyperText Markup Language (XHTML)

XHTML (del inglés, eXtensible HyperText Markup Language) es un lenguaje de programación pensado para sustituir a HTML. XHTML es la versión XML de HTML con las mismas funcionalidades, pero cumple las especificaciones más estrictas de XML.

### Java V1.6

Java es un lenguaje de programación orientado a objetos desarrollado por Sun Microsystems a principio de la década de 1990. El lenguaje en sí mismo toma gran parte de su sintaxis de C y C++, pero tiene un modelo de objetos más simple y elimina herramientas de bajo nivel, que suelen inducir a muchos errores, como la manipulación directa de punteros o memoria.

El lenguaje Java tiene la ventaja de ser multiplataforma. El mismo se ha extendido y ha cobrado cada día más importancia tanto en el ámbito de Internet como en la Informática en general. Propicia la integración de diversos frameworks que ofrecen múltiples facilidades tales como: persistencia, seguridad, logs, internacionalización, BPM, testing, web services, pantallas con elementos ricos de interfaz y relativa facilidad de uso.

### AJAX4JSF

AJAX4JSF es una librería open source o de código abierto que se integra totalmente en la arquitectura de JSF y extiende la funcionalidad de sus etiquetas, dotándolas con tecnología AJAX de forma limpia y sin añadir código JavaScript. Mediante esta librería se puede variar el ciclo de vida de una petición JSF, recargar determinados componentes de la página sin necesidad de recargarla por completo, realizar peticiones automáticas al servidor, controlar cualquier evento de usuario, entre otras funcionalidades. En definitiva AJAX4JSF permite dotar a una aplicación JSF de contenido mucho más ajustado a las necesidades del usuario, con muy poco esfuerzo.

### Java Persistence API (JPA)

JPA (del inglés, Java Persistence API), es la API de persistencia desarrollada para la plataforma Java en sus ediciones *Standard* (Java SE) y *Enterprise* (Java EE). Ha sido incluida en el estándar EJB 3.

La persistencia en este contexto cubre tres áreas:

- La API en sí misma, definida en `javax.persistence.package`
- La JPQL (del inglés, Java Persistence Query Language)
- Metadatos objeto/relacional

### Hibernate V3.3

Hibernate es una herramienta ORM (*Object-Relational Mapping* por sus siglas en inglés) para la plataforma Java que facilita el mapeo de atributos entre una base de datos relacional tradicional y el modelo de objetos de una aplicación, mediante archivos declarativos XML o anotaciones en los *beans* de las entidades que permiten establecer estas relaciones.

### JBoss Seam V2.1.1

JBoss Seam 2.1.1 es un potente framework para desarrollar aplicaciones web 2.0 al unificar e integrar tecnologías como AJAX, JSF, EJB, Java Portlets y BPM (del inglés, *Business Process Management*). Otra característica importante es que se pueden hacer validaciones en los POJOs (del inglés, *Plain Old Java Objects*) además de manejar directamente la lógica de la aplicación y de negocio desde las *session beans*.

### Facelets V1.1

Facelets es un framework simplificado de presentación, en donde es posible diseñar de forma libre una página web y luego asociarle los componentes JSF específicos. Aporta mayor libertad al diseñador y mejora los informes de errores que tiene JSF. Permite la definición de disposición de páginas basada en plantillas, la composición de componentes, creación de etiquetas personalizadas, desarrollo amigable para el diseñador gráfico y creación de librerías de componentes.

### Asynchronous JavaScript And XML

AJAX (del inglés, Asynchronous JavaScript And), no es un lenguaje de programación sino un conjunto de tecnologías tales como HTML, JavaScript, CSS (del inglés, *Cascading Style*), DHTML (del inglés, *Dynamic HTML*), PHP (del inglés, *HyperText Preprocessor*), ASP.NET (del inglés, *Application Service Providers*), JSP y XML que permiten hacer páginas de Internet más interactivas. La característica fundamental de AJAX es que permite actualizar parte de una página con información que se encuentra en el servidor sin tener que refrescar completamente la página. De modo similar su puede enviar información al servidor.

Ventajas en el uso de AJAX:

- Utiliza tecnologías ya existentes.
- Soportada por la mayoría de los navegadores modernos.
- Interactividad. El usuario no tiene que

esperar hasta que lleguen los datos del servidor.

- Portabilidad. No requiere plugins como Flash y Applet de Java.
- Mayor velocidad debido a que no hay que retornar toda la página nuevamente.
- La página se asemeja a una aplicación de escritorio.

Desventajas en el uso de AJAX:

- En algunos casos es necesario incluir señales para que el usuario se percate que el servidor ha respondido.
- AJAX depende de JavaScript y el cliente puede tener desactivada la opción de ejecutar el código JavaScript.
- Explotar una aplicación con excesivos recursos AJAX puede degradar el rendimiento del sistema.
- Una mala elección de tratar múltiples conexiones asíncronas puede ocasionar problemas.

### 2.2.2 Descripción del Algoritmo utilizado.

K-means fue creado por MacQueen<sup>1</sup> en 1967 y según [12] es uno de los algoritmos más simples de aprendizaje no supervisado, que resuelven el problema de la agrupación conocida. Sigue una forma fácil y simple para dividir una base de datos dada en *k* grupos (fijados a priori).

Simple K-Means

Para obtener un modelo no supervisado usando agrupamientos, se realizará utilizando el algoritmo Simple K-Means, que pertenece al grupo de algoritmos de partición-optimización. El algoritmo K-Means recibe como parámetro de entrada "*k*" y procede a dividir *n* objetos en "*k*" grupos, garantizando que los objetos de un mismo grupo sean semejantes entre sí y a su vez diferentes a los objetos de otros grupos. La similitud entre los grupos se mide desde el punto medio de los grupos, que puede ser visto como el centro de gravedad de los grupos. El objetivo de este método es crear grupos homogéneos en su interior y heterogéneos entre sí. Un criterio para evaluar la homogeneidad-heterogeneidad entre objetos es por la proximidad media de cada individuo del grupo. Esta puede ser determinada por la suma de los cuadrados de la diferencia de cada objeto con la media de cada grupo *j*. Esta función es conocida como la función objetivo.

La función objetivo más frecuentemente usada en técnicas de agrupamiento por partición es el error

---

<sup>1</sup> James B. MacQueen, profesor emérito de la Universidad de California, Los Ángeles.

cuadrático (*Squared Error* (SE)), que generalmente funciona bien con grupos compactos y bien separados. El error cuadrático de un agrupamiento formado por  $k$  grupos se expresa mediante la fórmula:

$$SE = \sum_{j=1}^k \sum_{i=1}^{N_j} \|x_i^j - c_j\|^2$$

**Figura 3. Fórmula de Error Cuadrático de la técnica de agrupamiento [13].**

Donde  $X_{ji}$  y  $C_j$  son el  $i$ -ésimo patrón y el centroide del  $j$ -ésimo grupo, respectivamente.

Teniendo como entradas:

$X \longrightarrow$  conjunto de datos.

$K \longrightarrow$  número de grupos.

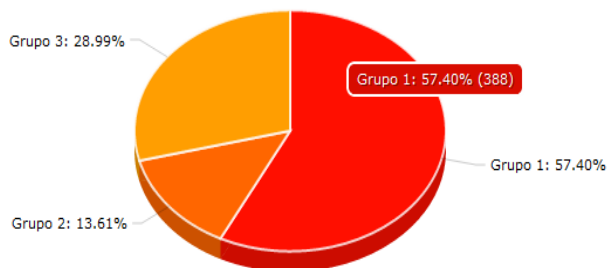
Este algoritmo fue seleccionado por las ventajas que presenta:

- Velocidad, la cual puede ser considerable cuando se trata de grandes volúmenes de datos.
- Buenos resultados.
- Posibilidad de cambiar los puntos iniciales y obtener resultados diferentes [14].

### 3. RESULTADOS Y DISCUSIÓN

A continuación se muestra el modelo obtenido después de haber aplicado el algoritmo Simple K-Means sobre los datos de entrenamiento almacenados en la tabla vista\_minable\_weka\_skm extraída del Almacén de Datos del SIAPS.

Antes de realizar un análisis a profundidad sobre este modelo, primero es necesario observar las características de cada grupo obtenido una vez aplicado el algoritmo para este caso.



**Figura 4. Datos procesados en 3 grupos.**

A partir de la interpretación conjunta de los modelos arrojados para el presente experimento se puede descubrir en el conjunto de datos lo siguiente:

- Grupo 1 (57.40%): se destacan las personas que se encuentran entre 45 y 65 años de edad, La distribución de los pacientes que tienen antecedentes patológicos familiares de Hipertensión Arterial es bastante uniforme, sin embargo se puede apreciar una mayoría de personas que tienen este tipo de antecedente sobre todo en las que se encuentran en el rango de edad previamente descrito con un 36.60%.
- Grupo 2 (13.61%): muy concentrado por personas de más de 65 años de edad. En este grupo, existe una distribución alta de casos de antecedentes familiares de Hipertensión Arterial, destacándose en esta edad un 70.65% del total de personas que padecen la enfermedad.
- Grupo 3 (28.99%): representa en su mayoría a las personas que son menores de 45 años de edad. Se puede apreciar una notable concentración de personas que si presentan antecedentes de Hipertensión Arterial en su familia, así como un 55.61% que padecen la enfermedad.
- Se puede apreciar que en los 3 grupos la generalidad de los pacientes que se encuentran agrupados son personas que tienen Hipertensión Arterial.

Una vez analizados los resultados obtenidos en la vista de análisis, se procedió a comprobar la información procesada utilizando la herramienta WEKA haciendo uso de la ventaja del presente sistema de generar y utilizar ficheros compatibles con dicha herramienta, los gráficos generados por esta aplicación internacionalmente reconocida en el procesamiento de datos arrojó altos niveles de concordancia en la exactitud de los resultados, afirmándose que la solución propuesta establece valores muy similares a los obtenidos en WEKA.

El sistema da la posibilidad de analizar los datos procesados a partir del informe estadístico que se genera tras la ejecución del algoritmo Simple K-Means:

Tiempo demorado en la ejecución: 0.064 seg  
Total de instancias procesadas: 676  
Número de iteraciones realizadas: 6  
Error cuadrático de grupos sin sumar: 3229.46495280981  
Valores perdidos globalmente reemplazados con media/modo

Tabla de grupos:

Grupos creados				
Atributos	Total (676)	Grupo 1 (282)	Grupo 2 (128)	Grupo 3 (274)
id_paciente	482.568	276.4965	738.9417	582.3759
rango_edad	Rango1	Rango2	Rango1	Rango1
genero_paciente	M	M	M	F
etnia_paciente	Negra	Mestiza	Negra	Negra
APP_ER	No	No	No	No
APP_EE	No	No	No	No
APP_EC	No	No	No	No
APP_HTA	No	No	Sí	Sí
APP_EC	No	No	No	No
APP_ER	No	No	No	No
APP_DM	No	No	No	No
CEF	No	No	Sí	No
DIS	No	No	No	No
PAL	No	No	No	No
EDE	No	No	No	No
DA	No	No	No	No
NAU	No	No	No	No
GR	No	No	No	No
EF	No	No	No	No
CT	No	No	Sí	No
AL	No	No	No	No
CA	No	No	No	No
DR	No	No	No	No
S	No	No	No	No
PAR	No	No	No	No
VER	No	No	No	No
DT	No	No	No	No
AV	No	No	No	No
DIA	No	No	Sí	No
IC	No	No	No	No
HTA	Sí	No	Sí	Sí

Figura 5. Informe estadístico del agrupamiento

#### 4. APOORTE Y NOVEDAD

El componente web desarrollado agiliza el proceso de análisis de la información. Mejora y apoya la interacción entre los diferentes especialistas para la toma de decisiones clínicas. Permite la comprensión de los datos procesados a través de gráficos y patrones de comportamiento. Dotará al Sistema Integral para la Atención Primaria de Salud de un soporte para la toma de decisiones que lo convertirá en un sistema más robusto, aportando así una herramienta para continuar perfeccionando el Sistema Nacional de Salud en Cuba, hoy de los más admirables del mundo.

#### 5. CONCLUSIONES

Los diferentes sistemas informáticos existentes en el mundo no satisfacen las necesidades actuales, por lo que se demostró la necesidad de desarrollar un componente web para apoyar la toma de decisiones clínicas. La Minería de Datos es un proceso eficaz para dar respuestas a preguntas complejas de Inteligencia de Negocios. Es una buena manera de convertir datos en información, y esta a su vez en conocimiento, para la correcta toma de decisiones clínicas. Como primeros pasos del desarrollo se analizaron las características del sistema, los méto-

dos a utilizar, las discusiones necesarias para la selección de las herramientas, tecnologías, lenguajes y la metodología que se utilizaría, manteniendo las políticas del departamento como base para esta selección. El algoritmo seleccionado es el correcto por tener ventajas como la velocidad, la cual puede ser considerable cuando se trata de grandes volúmenes de datos. Se obtienen buenos resultados tras la ejecución del mismo. Fue posible realizar una validación de los resultados obtenidos en la investigación corroborando la eficacia de la solución propuesta.

#### 6. REFERENCIAS BIBLIOGRÁFICAS

1. **Orallo Hernández, José, Ramírez Quintana, José Manuel and Ferri Ramírez, Cèsar.** INTRODUCCIÓN A LA MINERÍA DE DATOS. S.I.: Pearson, 2004.
2. **Orallo Hernández, Ref. 1.**
3. **Xu, Rui y Wunsch, Donald C.** Clustering. New Jersey: IEEE Press, 2009).
4. **Colectivo de autores.** Curso Universitario. Sistema de Información en los Sistemas de Salud. Italia: Instituto Universitario del Hospital Italiano.
5. **Pautsch, J.G.A.,** minería de datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación., Facultad de Ciencias Exactas, Químicas y Naturales. 2009, Universidad Nacional de Misiones: Argentina.
6. **Pautsch, J.G.A.** Ref. 5.
7. **Pautsch, J.G.A.** Ref. 5.
8. **Bressán, G.E.,** Almacenes de datos y minería de datos, T.y.s. distribuidos, Editor. 2003, Universidad Nacional del Nordeste. Facultad de Ciencias Exactas y Naturales y Agrimensura: Buenos Aires, Argentina.
9. **Fayyad, U. M., Piatetsky-Shapiro, G., Smith, P., Uthurusamy.** Advances in Knowledge Discovery and Data-Mining. s.l.: AAAI Press / The MIT Press, 1996.
10. **Frank, Ian H. Witten and Eibe.** Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann. 2000.
11. **Aplicación de minería de datos para el diagnóstico de accidentes cerebrovasculares agudos (ACVAs).** [En línea] [Citado el: 20 de enero de 2012.] [http://www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/DAEDALUS-MD19-Accidentes\\_Cardiovasculares.pdf](http://www.daedalus.es/fileadmin/daedalus/doc/MineriaDeDatos/DAEDALUS-MD19-Accidentes_Cardiovasculares.pdf).
12. **A Tutorial on Clustering Algorithms** [En línea] [Citado el: 6 de febrero de 2012.] [http://home.dei.polimi.it/matteucc/Clustering/tutorial\\_html/kmeans.html](http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html).

**13. A Tutorial on Clustering Algorithms,**  
Ref.12.

**14. DAEDALUS. [Online]** [Cited: febrero 7, 2012.] <http://www.daedalus.es>.

## **7. SÍNTESIS CURRICULARES DE LOS AUTORES**

**Alexeis Ochoa:** Nació en Holguín, Cuba, el 12 de noviembre de 1987, estudió en el Instituto politécnico de informática "Calixto García Iñiguez" de la ciudad de Holguín, con fecha de gradua-

ción junio de 2006, posteriormente ingresó en la Universidad de las Ciencias Informáticas(UCI) graduándose en esta de Ingeniero en Ciencias Informáticas en junio de 2012. Actualmente labora en la UCI, miembro del proyecto Sistema Integral para la Atención Primaria de Salud (SIAPS), del centro de desarrollo para soluciones de informática médica (CESIM). Desde 2011 es miembro de la asociación cubana de reconocimiento de patrones, cuentas con varios premios relevantes en ediciones de las Jornadas Científicas Estudiantil de la UCI y con una publicación en la serie científica de la Universidad de las Ciencias Informáticas (*Numero 5, Volumen 10 de 2012*). Su correo electrónico es [ajochoa@uci.cu](mailto:ajochoa@uci.cu).