



Descubrimiento de conocimiento en bases de datos históricas de una empresa comercializadora

Knowledge discovery in databases historical of a trading company

Tania Morgado-García^I, Daniel Antonio Ponce-de-León-Lima^{II}, Alejandro Rosete-Suárez^{III}

^I Corporación Copextel S.A., La Habana, Cuba

^{II} Universidad Estatal Península de Santa Elena, Santa Elena, Ecuador

^{III} Universidad Tecnológica de La Habana José Antonio Echeverría, La Habana, Cuba

Recibido: 16/09/2016

Aprobado: 22/09/2017

RESUMEN

Con el desarrollo de las tecnologías, el volumen ingente de las bases de datos y la necesidad de obtener conocimiento útil de ellas, la minería de datos se convierte en una herramienta estratégica para las empresas, como soporte a la toma de decisiones. Una empresa cubana creada en el año 1991, cuyo nombre no se menciona por razones de confidencialidad, la cual se distingue por ser un proveedor de soluciones integrales, ofrece productos y servicios ingenieros en una variada gama de esferas contando con un personal altamente capacitado. Posee grandes volúmenes de datos resultantes de la gestión de cada una de sus unidades organizativas, pero no se realizan análisis para identificar patrones y comportamientos ocultos en los datos. En este trabajo se desarrolló un proceso de Descubrimiento de Conocimiento en Bases de Datos en la mencionada empresa como ayuda a la toma de decisiones. En este sentido, se detallaron aspectos relacionados con la Minería de Datos y su aplicación en el entorno de los indicadores económicos-financieros y de gestión, se obtuvieron modelos de asociación y agrupamiento que permitirán apoyar el trabajo de los directivos en los diferentes niveles de la organización.

Palabras clave: minería de datos, agrupamiento, conocimiento, empresa.

ABSTRACT

With the development of technologies, the big volume of databases and the need to obtain useful knowledge of them, data mining becomes a strategic tool for companies, which support the decision making. A cuban company, created in 1991, whose name is not mentioned for reasons of confidentiality, which is distinguished as a provider of comprehensive solutions, offers products and engineering services in a wide range of areas with a highly trained staff. It has large volumes of data resulting from the management of each of its organizational units, but no analysis are done to identify hidden patterns and behaviors in data. This paper present a process of Knowledge Discovery in Databases in the aforementioned company as an aid to decision making. In this regard, aspects related to data mining and its application in the environment of economic-financial indicators and management were detailed, association and clustering models that will support the work of managers at different levels of the organization were obtained.

Key words: data minig, grouping, knowledge, company.

I. INTRODUCCIÓN

En los últimos años, el desarrollo tecnológico tanto en el área de cómputo como en la de transmisión de datos, ha hecho posible que se gestionen de una mejor manera el almacenamiento y manejo de grandes volúmenes de datos. Las empresas de hoy se mueven en entornos altamente competitivos y de cambio continuo. La dinámica del mercado conduce a la necesidad de contar con la información adecuada en el momento indicado para que los directivos puedan tomar las decisiones de negocio apropiadas. Por ello, han comprendido que los grandes volúmenes de datos que residen en sus sistemas pueden, y deben, ser analizados y explotados para obtener nuevo conocimiento [1], [2].

La Minería de Datos o Explotación de Información, es el proceso de extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos, siendo su principal objetivo encontrar información oculta o implícita, que no es posible obtener mediante métodos estadísticos convencionales. La entrada al proceso de minería está formada por contenedores de información diversos, esto incluye bases de datos relacionales, almacenes de datos (Datawarehouse), documentos en texto libre, datos de la Web, entre otros [3], [4].

Una definición tradicional la definen en [5], como un paso del proceso de KDD (Knowledge Discovery in Databases) que consiste en aplicar análisis a los datos y descubrir algoritmos que, bajo limitaciones de eficiencia computacional aceptables, producen un grupo particular de patrones o modelos sobre los datos.

La minería de datos ofrece la oportunidad a las organizaciones de detectar problemas potenciales y tomar las acciones para evitarlos o mitigar el impacto. Aplicada apropiadamente, la Minería de Datos le reporta a las organizaciones numerosos beneficios entre los que se encuentran [6]: mayor eficiencia a partir de que se consume menos tiempo en corregir errores y en recuperar pagos; información anticipada que permite reducir las sorpresas y analizar las causas de los problemas, y mejores controles a través de la prevención y detección de actividades posibilitando el monitoreo de huecos de control interno.

Su campo de aplicación es amplio y diverso, de hecho, en todas las actividades donde se generen datos. En [7], [8] y [9] se exponen varios ejemplos de su aplicación.

Con un enfoque no tradicional, en [10] los autores presentan la minería de datos como un proceso sistemático y social que involucra los diferentes sistemas y fases de la minería así como sus respectivos entornos. En la minería de datos, los problemas y aplicaciones están normalmente rodeados de factores organizacionales y sociales en las diferentes empresas. Tales factores conforman el entorno de minería del mundo real y se hace necesaria su atención, ya que reflejan las necesidades reales de la empresa y los usuarios, y constituyen el entorno de los conocimientos identificados. La inteligencia que surge de la participación de factores y recursos de la organización en los patrones de minería constituye la inteligencia organizacional. Consta de muchos aspectos entre los que se encuentran [10]:

- Las estructuras organizativas relacionadas con aspectos claves, como de dónde provienen los datos.
- El comportamiento organizacional, en cuanto a la comprensión del negocio y los datos, cómo los individuos y grupos actúan en la organización.
- Las regulaciones del negocio que incluyen reglas, políticas, protocolos, normas.
- Los procesos de negocio y los flujos de trabajo.
- Los actores de la organización y sus funciones.

El proceso de KDD

Fayyad et al. en [5] definen el proceso de KDD como: "Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos". Está compuesto por cinco fases, las cuales se describen a continuación [11]:

1. Integración y recopilación de datos: En esta fase se determinan las fuentes de información que pueden ser útiles y dónde conseguir las.
2. Selección, limpieza y transformación: Incluye detectar los outliers, los datos faltantes o perdidos, construir nuevos atributos y numerizar o discretizar los atributos.
3. **Minería de datos:** El objetivo de esta fase es producir nuevo conocimiento. Para ello se construye un modelo que no es más que una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones pasadas. Incluye:
 - Elegir el tipo de modelo.
 - Determinar qué tipo de tarea de Minería es el más apropiado.

DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS HISTÓRICAS DE UNA EMPRESA COMERCIALIZADORA

- Elegir el algoritmo de Minería que resuelva la tarea y obtenga el tipo de modelo que se está buscando.
4. Evaluación e interpretación: Existen diferentes medidas de evaluación de los modelos: precisos, comprensibles (inteligibles) e interesantes (útiles y novedosos).
 5. Difusión: Después que el modelo es construido y validado, este puede utilizarse en disímiles finalidades, y durante este proceso deben medirse su evolución. De esta manera se sabrá si debe ser reevaluado, re-entrenado y posiblemente reconstruido completamente.

Son diversos los elementos a tener en cuenta para que una organización implemente con éxito el proceso de descubrimiento de conocimiento o minería de datos. En [6] los autores definen un conjunto de aspectos que, más que elementos técnicos, representan cambios organizativos, de mentalidad y culturales en las organizaciones:

- La aceptación de la responsabilidad para evitar los problemas de forma proactiva mediante la adopción de enfoques y herramientas de minería de datos. Esto es esencialmente una transformación cultural de la organización.
- Comprender el potencial apoyo que representa la minería de datos a la organización tanto en la gestión diaria como en la toma de decisiones estratégicas.
- Determinar cómo se va a utilizar el conocimiento resultante del proceso de minería de datos.
- Compartir experiencias y mejores prácticas.
- Ver este proceso como un proceso continuo.

Las actividades de minería de datos se clasifican en unas pocas categorías que dependen de las tareas y los objetivos del análisis. Dependiendo de la existencia de una variable objetivo, el proceso de aprendizaje se clasifica en [3]:

- Supervisado o MD directa. Un atributo objetivo representa la clase a la cual cada registro pertenece o expresa una cantidad medible. Se orientan a la predicción e interpretación respecto a un atributo objetivo.
- No supervisado o MD Indirecta. Los análisis no están determinados por un atributo objetivo. Su propósito es descubrir patrones recurrentes y afinidades en conjuntos de datos.

A partir del propósito de la minería de datos indirecta, resulta conveniente su aplicación en las organizaciones que no poseen una historia previa de uso de la misma.

Una empresa cubana creada en el año 1991 que se distingue por ser un proveedor de soluciones integrales, ofrece productos y servicios ingenieros en una variada gama de esferas, contando con un personal altamente capacitado [12].

Gran parte de la gestión de la empresa se encuentra soportada sobre las tecnologías de información y las comunicaciones. Existen un conjunto de aplicaciones informáticas que soportan los procesos claves, de apoyo y de dirección, a través de las cuales se registran en cada unidad organizativa los datos primarios resultantes de su actividad. Estos datos se replican a los servidores centrales en los que se dispone de grandes volúmenes que datan del año 2001. Entre las aplicaciones se destacan las que permiten realizar de forma automatizada el proceso de cierre mensual de la contabilidad, y la creación de indicadores económicos y reglas que validan la información contable a partir de los indicadores calculados. Sin embargo, no se aprovecha el conocimiento oculto en estos datos, que pueda sustentar determinadas acciones y decisiones por parte de los directivos de la empresa.

Del problema anterior se deriva el objetivo del presente trabajo: descubrir patrones, tendencias y desviaciones en el almacén de datos corporativo de la empresa aplicando técnicas no supervisadas de minería de datos con vistas a mejorar el proceso de toma de decisiones. Los datos seleccionados para inicialmente explorar y buscar patrones y comportamientos ocultos son los contables, específicamente los indicadores económicos calculados a partir de la aplicación que lleva el mismo nombre.

De esta forma se pretende determinar las relaciones que existen entre los diferentes indicadores económicos que reflejen patrones y/o comportamientos, y agrupar las unidades organizativas atendiendo al comportamiento de estos indicadores.

II. MÉTODOS

A continuación, se describen las tareas de minería de datos realizadas, metodología y herramienta de análisis de datos usadas, así como las técnicas y algoritmos de asociación y agrupamiento utilizados.

Reglas de Asociación

La primera tarea de minería de datos realizada fue la obtención de reglas de asociación la cual consiste en identificar patrones regulares y recurrentes dentro de un gran grupo de transacciones. Las reglas aparecen de muchas formas pero la formulación más usual es del estilo "si el atributo X toma el valor d entonces el atributo Y toma el valor b". Se evalúan usando dos parámetros: *confianza* y *soporte* que también se nombran precisión y cobertura respectivamente [13].

Agrupamiento

La segunda tarea fue el agrupamiento la cual tiene como objetivo subdividir los registros de un grupo de datos en grupos homogéneos de observaciones, llamados Clusters. Los datos se agrupan basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los distintos grupos [11].

Metodología y herramienta de análisis de datos

Se decidió emplear la metodología CRISP-DM[14] para el desarrollo de esta investigación porque mantiene una perspectiva más amplia respecto a los objetivos empresariales y concibe el proyecto de KDD de forma global. Además, es una metodología neutra respecto a la herramienta que se utilice para el desarrollo del trabajo y su distribución es libre y gratuita. En la fase de minería de datos se utilizó como método para encontrar las asociaciones el algoritmo **A priori**[3], [15], y para encontrar los grupos el algoritmo **K-Medias**[3], [16], ambos desarrollados en la herramienta de análisis de datos **R**[17], [16].

III. RESULTADOS Y DISCUSIÓN

Fase de preparación de los datos

Los datos utilizados en la presente investigación proceden de la tabla *H_Indicadores* en la que se almacenan, después de calculados a partir de la información contable que se replica a los servidores centrales de cada una de las unidades organizativas de la empresa, los indicadores económicos resultantes de su operación y gestión. Se modificó, mediante consulta SQL, la estructura original del conjunto de datos de forma tal que cada uno de los indicadores constituya una variable, y de esta forma facilitar su análisis. Se eliminaron registros con valores nulos. La estructura resultante se muestra en la Tabla 1:

La selección de características reduce el tamaño de los datos eligiendo las variables más influyentes en el problema. La revisión detallada de cada indicador refleja que existen indicadores que totalizan, es decir que son el resultado de operaciones aritméticas entre otros indicadores, o unos contienen a otros. Además, todos los indicadores son variables contables de la empresa. Por este motivo se decide calcular nuevos indicadores de eficiencia que permitan realizar análisis y comparaciones a partir de las variables originales.

Tabla 1. Estructura modificada de conjunto de datos

Tabla	Campos	Descripción
H_Indicadores modificado	Id_gerencia	id de la división
	Nombre_Gerencia	Nombre de la División
	id_sucursal	id de la sucursal que pertenece para esa división
	Nombre_Sucursal	Nombre de la Sucursal
	id_moneda	id de la moneda en la que se calculó el indicador
	id_fecha	id de la fecha que se corresponde con el mes y año para ese indicador
	Mes	
	año	
	Indicador1	
	Indicador2	
	.	
	Indicador N	

Construcción de características

La creación o construcción de características consiste en crear nuevos atributos para mejorar la calidad, visualización y comprensibilidad del conocimiento extraído.

En la Tabla 2 se muestran los nuevos indicadores calculados a partir de los originales, con una breve descripción y fórmula de cálculo.

DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS HISTÓRICAS DE UNA EMPRESA COMERCIALIZADORA

Tabla 2. Nuevos Indicadores calculados

Código	Descripción	Expresión de Cálculo
mub	Margen de Utilidad Bruta	Utilidad Bruta/Ventas Netas*100
muo	Margen de Utilidad de Operaciones	Utilidad Operaciones/Ventas Netas*100
mn	Margen Neto	Utilidad Neta/Ventas Netas*100
rsa	Rentabilidad sobre Activos	Utilidad Operaciones/Total Activos*100
rc	Razón Corriente	Activo Circulante/Pasivo Circulante
fm	Fondo de Maniobra	Activo Circulante-Pasivo Circulante
rct	Rotación Capital de Trabajo	Ventas Netas/Fondo de Maniobra
cs	Cobertura de Stocks	Fondo de Maniobra/Inventario*100
rcv	Relación con las ventas	Fondo de Maniobra/Ventas Netas*100
re	Rotación de Existencias	Ventas Netas/Inventario
ra	Rotación de Activos	Ventas Netas/Total Activos
raf	Rotación de Activos Fijos	Ventas Netas/Activos Fijos
rend	Razón de endeudamiento	Total Pasivos/Total Activos*100
inm	Inmovilización	Activo Fijo/Total Activos*100
iac	Importancia del Activo Corriente	Activo Circulante/Total Activo*100

Con el objetivo de obtener datos con mayor calidad y cumplir los requerimientos de las técnicas seleccionadas se realizan nuevas transformaciones sobre las variables. Para ello se utiliza la técnica de discretización. Se segmentaron cada uno de los indicadores en 6 grupos atendiendo a los requerimientos y valores esperados de cada uno. Además, se discretizaron las variables IdGerencia según el tipo de unidad organizativa (División Comercial, División de Apoyo, Servicios Técnicos o División Territorial), y mes según los trimestres del año. En la Tabla 3 se muestran las variables resultantes y el método de segmentación utilizado.

Tabla 3. Variables resultantes para análisis.

Variable	Descripción	Método de segmentación
IdGerencia	Identificador de la División o Gerencia	Se agruparon las divisiones según su tipo (A-Comerciales, B-Territoriales, C-Servicios Técnicos y D-Apoyo)
NombreG	Nombre de la División o Gerencia	-
Ano	Año	-
Mes	Mes	Se discretiza atendiendo a los trimestres del año
mub	Margen de Utilidad Bruta	Mediante agrupación por K-Medias
muo	Margen de Utilidad de Operaciones	Mediante agrupación por K-Medias
mn	Margen Neto	Mediante agrupación por K-Medias
rsa	Rentabilidad sobre Activos	Mediante agrupación por K-Medias
rc	Razón Corriente	Se discretiza de forma manual según los criterios del indicador
fm	Fondo de Maniobra	Mediante agrupación por K-Medias
rct	Rotación Capital de Trabajo	Mediante agrupación por K-Medias
cs	Cobertura de Stocks	Mediante agrupación por K-Medias
rcv	Relación con las ventas	Se discretiza de forma manual según los criterios del indicador
re	Rotación de Existencias	Mediante agrupación por K-Medias
ra	Rotación de Activos	Mediante agrupación por K-Medias
raf	Rotación de Activos Fijos	Mediante agrupación por K-Medias
rend	Razón de endeudamiento	Se discretiza de forma manual según los criterios del indicador
inm	Inmovilización	Mediante agrupación por K-Medias
iac	Importancia del Activo Corriente	Mediante agrupación por K-Medias

Fase de Minería de datos: Tarea descriptiva Reglas de asociación

Se utilizó el algoritmo A priori. Se obtuvieron valores de confianza y soporte muy elevados. Se encontraron reglas con valores de soporte y confianza alrededor de 0.80 y 0.90 respectivamente. A continuación, las reglas de mayor interés:

- La cobertura de stock en el rango [0, 447], el margen de utilidad de operaciones en el rango [0, 46.6] % y la rotación del capital de trabajo en el rango [0, 8.86] implican que la rotación de existencias se encuentra en el rango [0, 11.9].
- La rotación de existencias en el rango [0, 11.9] implica que el margen de utilidad de operaciones se encuentra en el rango [0, 46.6] %, y viceversa.
- La razón de endeudamiento en el rango [60,122] % implica que la rotación de existencias se encuentra en el rango [0,11.9].
- La razón de endeudamiento en el rango [60,122] % y la rotación de existencias en el rango [0,11.9] implican que el margen de utilidad de operaciones se encuentra en el rango [0,46.6] %.
- La razón corriente en el rango [0,1.5] y la razón de endeudamiento en el rango [60,122] % implican que el margen de utilidad de operaciones se encuentra en el rango [0,46.6] %.
- La cobertura de stocks en el rango [0,447] implica que la rotación de existencias se encuentra en el rango [0,11.9].
- La cobertura de stocks en el rango [0,447] implica que el margen de utilidad de operaciones se encuentra en el rango [0,46.6] %.
- El margen de utilidad de operaciones en el rango [0, 46.6] % y la razón de endeudamiento en el rango [60, 122] % implican que la rotación de existencias se encuentra en el rango [0,11.9].
- La rotación del capital de trabajo en el rango [0,8.86] implica que la rotación de existencias se encuentra en el rango [0,11.9].
- Si la división es comercial (A) implica que la rotación de existencias se encuentra en el rango [0,11.9].

Todos los indicadores se utilizaron en el algoritmo aplicado, pero algunos no resultaron interesantes: margen de utilidad bruta, margen neto, rentabilidad sobre activos, fondo de maniobra, rotación de activos y rotación de activos fijos, entre otros. Por otra parte, como se aprecia en las reglas anteriormente expuestas, las características que mejor describen las relaciones, son: cobertura de stock, margen de utilidad de operaciones, rotación del capital de trabajo, rotación de existencias, razón de endeudamiento y razón corriente. Los resultados obtenidos muestran las relaciones entre los indicadores, algunas lógicas y evidentes pero que ofrecen otra perspectiva para el análisis de la actividad en la empresa. Representan un conocimiento valioso para la organización al disponer de los indicadores recurrentes, las asociaciones entre ellos y en que rango de valores se encuentran.

En este sentido, destacar que la razón de endeudamiento en todos los casos se presenta en el rango [60, 122] %. El valor recomendado se ubica entre 40 y 60%, no obstante, cuando tiende a cero la empresa tiene más solidez porque se financia con recursos propios. En nuestro caso, se encuentra corrido el valor por encima del límite superior recomendado.

Fase de Minería de datos: Tarea descriptiva Agrupamiento

En esta sección se muestran los experimentos realizados con el objetivo de encontrar posibles agrupaciones entre las instancias acorde a sus características, aplicando el algoritmo K-Medias en R.

Para lograr mayor precisión en la ejecución del algoritmo se eliminan de la vista minable inicial las observaciones con valores NA en cualquiera de sus atributos. Por este motivo la cantidad de instancias a utilizar se reduce 5453.

Para determinar la cantidad de grupos ideal se ejecutó la técnica Suma de Cuadrados de Error o SSE (Sum of Squares Error en inglés). Se realizaron análisis de grupos anidados desechando los grupos con baja representatividad (inferior a 1%).

Descripción del agrupamiento

Se pudo observar la conformación de 4 grupos con las siguientes distribuciones:

- Grupo 1 (499 observaciones que representan el 9.29 %)
- Grupo 2 (1260 observaciones que representan el 23.45 %)
- Grupo 3 (400 observación que representa el 7.44 %).
- Grupo 4 (3214 observaciones que representan el 59.82 %)

DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS HISTÓRICAS DE UNA EMPRESA COMERCIALIZADORA

El **Grupo 1** se caracteriza por:

- Valores mayores del fondo de maniobra lo que significa que se dispone de activos circulantes suficientes para cubrir las deudas a corto plazo.
- Valores menores de inmovilización de activos, que significa porcentos menores de capital inmovilizado.
- Valores mayores de margen de utilidad bruta. El valor deseado para este indicador es el mayor posible e indica la proporción del ingreso que permite cubrir los gastos diferentes al costo de venta.
- Valores mayores de margen neto lo que significa mayores beneficios y mayor rentabilidad.
- Valores menores de rotación de activos lo que pudiera indicar sobredimensionamiento o capacidad ociosa en las inversiones. El valor deseado para este indicador es el mayor posible. Cuanto mayor es su valor, en mayor medida la empresa utiliza sus recursos para generar negocio.

El **Grupo 2** se caracteriza por:

- Valores mayores de razón corriente lo que pudiera significar que existen recursos ociosos.
- Valores mayores de rotación del capital de trabajo. Esto significa mejor productividad del capital de trabajo, lo cual se traduce en una mejor rentabilidad. Se debe tener presente que un índice elevado puede revelar insuficiencia del fondo de maniobra.
- Valores de relación con las ventas en el mayor rango, lo que significa que en este grupo se encuentran las observaciones en las que el porcentaje de las ventas que corresponde al fondo de maniobra es mayor. El valor adecuado para este indicador es entre el 15 y 20%.

El **Grupo 3** se caracteriza por:

- Valores mayores de inmovilización de activos, que significa mayores porcentos de capital inmovilizado.
- Valores negativos de cobertura de stocks lo que significa que el fondo de maniobra no cubre los inventarios.
- Valores negativos de fondo de maniobra lo que indica que no se cuenta con suficientes activos para hacer frente a las deudas a corto plazo.
- Valores negativos de margen neto, rotación de capital de trabajo y relación con las ventas. Los valores de estos tres indicadores deben ser mayores que cero. En el caso de la relación con las ventas el valor razonable del indicador es entre el 15 y el 20%. Se debe tener en cuenta que estos valores negativos pueden corresponder a contabilizaciones o registros incorrectos.
- Valores mayores de rentabilidad sobre activos. El valor deseado para este indicador es el mayor posible y mide las utilidades que generan los recursos totales con que cuenta el negocio.
- Valores mayores de razón de endeudamiento. El valor recomendado para este indicador es entre el 40 y 60 %.

El **Grupo 4** se caracteriza por:

- Valores negativos de margen neto. El valor de este indicador debe ser mayor que cero.
- Valores mayores de margen de utilidad de operaciones. El valor deseado para este indicador es el mayor posible e indica que se está generando suficiente utilidad para cubrir las necesidades de financiamiento.
- Valores negativos de rotación de capital de trabajo. El valor de este indicador debe ser mayor que cero.

En los grupos 1, 2 y 3 los valores del indicador Razón de Endeudamiento presentan un comportamiento similar, pero es necesario destacar que se encuentran en un rango superior al deseable para este indicador que es entre el 40 y 60%. Los valores de los indicadores importancia del activo corriente, rotación de activos fijos y rotación de existencias presentan un comportamiento similar en todos los grupos, y se encuentran en rangos razonables para estos indicadores.

Los valores de los indicadores cobertura de stocks y rotación de activos fijos presentan un comportamiento similar en los grupos 1, 2 y 4, y se encuentran en rangos razonables para estos indicadores. De manera similar, los valores del indicador margen de utilidad de operaciones presentan un comportamiento similar en los grupos 1, 2 y 3, y se encuentra en un rango razonable para este indicador.

De lo anterior se concluye que en el **Grupo 1** se concentran las observaciones resultantes de la actividad de cada una de las unidades organizativas con mejor desempeño, atendiendo a los valores promedio de cada uno de los indicadores. Los años más representativos se encuentran

entre el 2009 y 2013 y las unidades organizativas se corresponden con divisiones comerciales que se repiten en estos años.

Por su parte, en el **Grupo 3** se concentran las observaciones con valores de indicadores alejados de los valores deseables o recomendados. Los años más representativos son 2005, 2008 y 2012. Las unidades organizativas comprenden divisiones territoriales, de apoyo y comerciales e igualmente se repiten en estos años.

En la Figura 1 se muestran (a modo de ilustración) las distribuciones de las observaciones de cada grupo en el tiempo. En el eje de las abscisas se presenta el tiempo por cada uno de los meses (de enero a diciembre) de los años representados, y en el de las ordenadas la cantidad de observaciones correspondientes a cada mes y año.

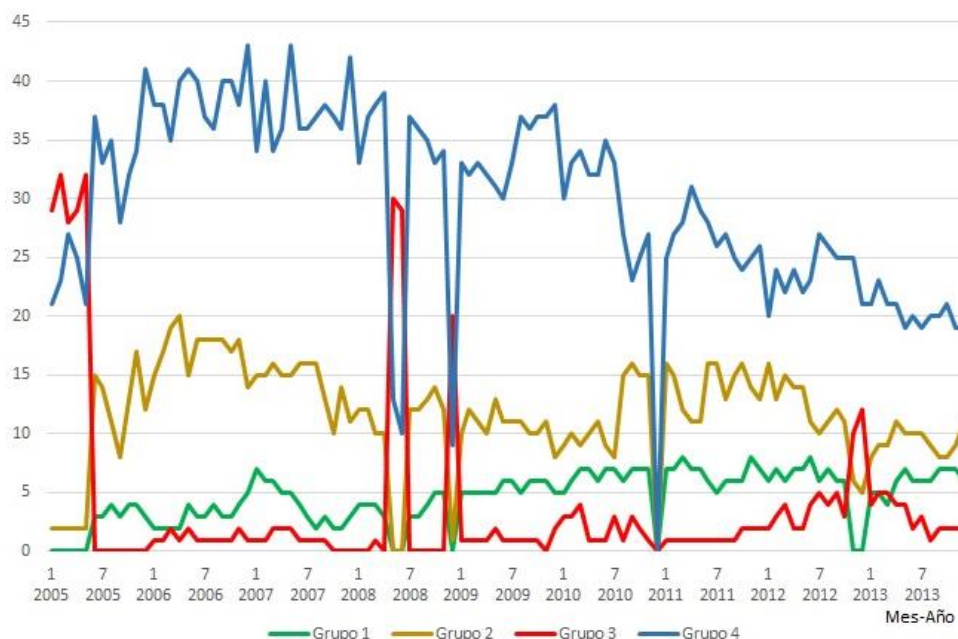



Fig.1. Distribución de frecuencia por grupos

Como se aprecia en la Figura 1, el Grupo 4 contiene la mayor cantidad de observaciones y se evidencia la presencia de las observaciones correspondientes al Grupo 3 con mayor incidencia en los años 2005, 2008 y 2012.

IV. CONCLUSIONES

La minería de datos como parte del proceso de KDD constituye una herramienta estratégica para las empresas y de apoyo en la toma de decisiones. Por su parte, la minería de datos indirecta permite descubrir patrones recurrentes y relaciones ocultas en conjuntos de datos, por lo que resulta conveniente su aplicación en las organizaciones que no poseen una historia previa de uso de la minería de datos. Se considera que se cumplieron los objetivos propuestos, se obtuvieron reglas que presentan parámetros aceptados para ser consideradas útiles durante la toma de decisiones por la dirección de la empresa. Igualmente, los grupos obtenidos (patrones) describen el comportamiento asociado a los valores de cada uno de los indicadores, y representan un apoyo al trabajo de funcionarios y directivos. 

VI. REFERENCIAS

1. Timarán Pereira R. Arquitecturas de integración del proceso de descubrimiento de conocimiento con sistemas de gestión de bases de datos: un estado del arte. *Ing Compet.* 2001;3(2):45-55.
2. Rygielski C, Wang JC, Yen DC. Data mining techniques for customer relationship management. *Technol Soc.* 2002;24(4):483-502.
3. Vercellis C. *Business Intelligence: Data Mining and Optimization for Decision Making*: John Wiley & Sons; 2009.
4. Berry MJA, Linoff G. *Data mining techniques: for marketing, sales, and customer relationship management*. 2nd ed. Indianapolis: Wiley Pub; 2004.
5. Fayyad U, Piatetsky Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Mag.* 1996;17(3):37.
6. Steinhoff JC, Carnahan TL. *Smart Use of Data Mining is Good Business and Good*

- Government. *J Gov Financ Manag.* 2012:17-22.
7. Vimal Kumar D, Tamilarasi A. Genetic algorithm-based optimized association rule mining for multi-relational data. *Intell Data Anal.* 2013;17:965-80.
 8. Lee BJ, Kim JY. Identification of Type 2 Diabetes Risk Factor using Phenotypes consisting of Anthropometry and Triglycerides based on Machine Learning. *IEEE J Biomed Health Inform.* 2015:1.
 9. Ting IH, Lin YC, et al. What is Missing? Using Data Mining Techniques with Business Cycle Phases for Predicting Company Financial Crises. *Asia Pac Manag Rev.* 2011;16(4):535-49.
 10. Cao L, Yu PS, Zhang C, et al. *Domain Driven Data Mining.* Boston, US: Springer 2010.
 11. Hernández Orallo J, Ramírez Quintana MJ, Ferri Ramírez C. *Introducción a la Minería de Datos:* Pearson Prentice Hall; 2004.
 12. Copextel. Copextel La Solución Integral. [Citado 15 de julio de 2016]. Disponible en: www.copextel.com.cu
 13. Witten IH, Frank E, Hall MA. *Data mining: practical machine learning tools and techniques.* 3rd ed. Burlington: Morgan Kaufmann; 2011. ISBN. DOI [Citado
 14. Chapman P, Clinton J, Kerber R, et al. *CRISP-DM 1.0 Step-by-step data mining guide.* 2000.
 15. Hahsler M, Buchta C, Gruen B, et al. *Arules: Mining Association Rules and Frequent Itemsets.* R package version 1.1-6. 2014.
 16. Core Team R. *A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2014.
 17. Venables WN, Smith R. Development Core Team, y others. En: *An introduction to R* versión 312 Network Theory Ltd; 2002.