

Estudio de la gestión del presupuesto en el ISMMM utilizando minería de datos^{*1}

[Study of budget management in ISMMM using data mining]

YIEZENIA ROSARIO FERRER², YANELIS GÉ GUILARTE³

RECIBO: 28.07.2013 – APROBACIÓN: 02.12.2013

Resumen

La elaboración del presupuesto permite a las empresas, los gobiernos e instituciones establecer prioridades y evaluar la consecución de sus objetivos. Su confección se basa en las instituciones de educación superior en sus procesos sustantivos. En los últimos años ha existido un gran crecimiento en las capacidades de generar y almacenar datos, de ahí que el volumen y la variedad de información hayan ido en aumento. Esto ha traído como consecuencia la incapacidad para analizar y transformar la información en conocimiento útil con las herramientas disponibles. En la presente investigación se realiza el descubrimiento de conocimiento oculto en los datos del presupuesto almacenados desde el 2004, registrados en documentos en formato Excel, por medio del uso de las técnicas de minería de datos: regresión lineal, asociación y clasificación; en la obtención de modelos para lograr mejoras en el proceso de elaboración del presupuesto en el ISMMM. El proceso de obtención de conocimiento fue realizado utilizando la metodología CRISP-DM y a la herramienta de análisis de datos WEKA.

Palabras clave: CRISP-DM, gestión de presupuesto, minería de datos, WEKA.

- * Modelo para la citación de este artículo:
ROSARIO FERRER, Yiezenia & GÉ GUILARTE, Yanelis (2013). Estudio de la gestión del presupuesto en el ISMMM utilizando minería de datos. En: Ventana Informática No. 29 (jul-dic). Manizales (Colombia): Facultad de Ciencias e Ingeniería, Universidad de Manizales. p. 81-93. ISSN: 0123-9678
- 1 Reporte de caso proveniente del proyecto *Herramientas informáticas de apoyo a la gestión presupuestaria en el ISMMM*, ejecutado en el periodo 09.2011-12.2012, e inscrito en el grupo de investigación *Inteligencia Artificial Aplicada* de la *Facultad Geología y Minería. Departamento Informática y Matemáticas del Instituto Superior Minero Metalúrgico de Moa (ISMMM)*.
 - 2 Licenciada en Ciencia de la Computación, Doctora en Informática. Profesor Auxiliar, Instituto Superior Minero Metalúrgico de Moa (Moa, Cuba). Correo electrónico: jessie@ismm.edu.cu
 - 3 Ingeniera Informática. Instructor, Instituto Superior Minero Metalúrgico de Moa (Moa, Cuba). Correo electrónico: yge@info.ismm.edu.cu

Abstract

The budget planning allows to enterprises, governments and institutions to states the priorities and to assess how its objectives are carried out. In higher education institutions, the budget planning is based on its main processes. In the last years it had existed an accelerated growing of the capacities to generate and store data, so the volume and variety of information also grown. That's why the disabilities to analyze and transform this information in useful knowledge with the available tools. This research aims to discover hidden knowledge from budget data stored in Excel files in the years from 2004 to 2011 using the data mining techniques: linear regression, association and classification, to improve the budget forecasting process at ISMMM. The knowledge discovering was made using the CRISP-DM methodology and the data analysis tool WEKA.

Keywords: Budget management, CRISP-DM, data mining, WEKA.

Introducción

El cálculo anticipado de los ingresos y gastos de una actividad económica durante un período se denomina presupuesto, es un plan dirigido a cumplir una meta prevista, expresada en valores y términos financieros. De ahí que elaborar el presupuesto permite a las empresas, los gobiernos e instituciones establecer prioridades y evaluar la consecución de sus objetivos.

Sweeny & Rachlin (1984, 2), definen que el presupuesto es la presentación ordenada de los resultados previstos de un plan, un proyecto o estrategia. Para Burbano (2005, 11) son una expresión cuantitativa formal de los objetivos que se propone alcanzar la administración de la empresa en un período, con la adopción de las estrategias necesarias para lograrlos.

De manera general, el presupuesto es un plan integrador expresado en términos financieros relativo a las operaciones y recursos que forman parte de una empresa para un período determinado, con el fin de lograr los objetivos fijados, según considera Nelson (2006, 49). Es función del presupuesto ser una guía de acción, lo que permite comparar lo real con lo planificado, optimizar el empleo económico de los recursos, así como trabajar en forma anticipada previniendo los hechos, aseguran Burbano & Ortiz (1995, 16). En consecuencia los presupuestos son importantes porque ayudan a minimizar el riesgo en las operaciones de la organización.

La Dirección de Economía del Instituto Superior Minero Metalúrgico de Moa, ISMMM, tiene entre sus tareas la elaboración del anteproyecto del

presupuesto, la aprobación y el control de la ejecución del presupuesto de la entidad. El anteproyecto del presupuesto está sustentado por una fundamentación teórica que explica las variaciones del presupuesto por los distintos programas. Su confección se basa en las proyecciones previstas para el año, teniendo como base el nivel de actividad y los objetivos propuestos. Para ello se consideran los procesos sustantivos: formación de pregrado y posgrado, ciencia e innovación tecnológica y extensión universitaria; y las actividades de apoyo, tales como administración, alimentación, mantenimiento y reparaciones, relaciones internacionales, entre otras.

Este proceso genera gran volumen de información, lo cual en las condiciones actuales puede considerarse como un activo principal de la organización. Sin embargo, esta información no es utilizada como referente en la elaboración del presupuesto, pues los especialistas encargados de esta tarea se basan fundamentalmente en su intuición y en el análisis de los porcentajes de ejecución de las partidas del presupuesto en el año anterior para realizar sus propuestas de anteproyecto.

Por tal razón, deben orientarse los recursos tecnológicos de manera que ayuden a tomar decisiones estratégicas y oportunas ante la ausencia de un modelo que facilite la toma de decisiones respecto al presupuesto asignado a las distintas partidas y elementos del presupuesto, así como cuál debe ser la asignación de presupuesto a las distintas áreas, facultades y departamentos del instituto. El conocimiento del comportamiento histórico de las partidas y elementos del presupuesto podría ser de suma importancia en la elaboración y desagregación del presupuesto para los años futuros.

La minería de datos como proceso de descubrir patrones útiles ha sido utilizada en la solución de problemas relacionados con el mundo de las finanzas. Este proceso iterativo cuyo progreso está definido por el descubrimiento, de acuerdo con Kirkos, Spathis & Manolopoulos (2007, 996), ha sido utilizado en la predicción de mercados, en la predicción de quiebra, en la detección de fraudes, entre otros.

Considerando la amplia aplicabilidad de la minería de datos en las finanzas, ha sido valorada su utilización para el desarrollo de esta investigación, por su capacidad para la generación de modelos tanto descriptivos como predictivos. Es importante mencionar, que los modelos descriptivos permiten comprender los datos e identificar patrones, relaciones y dependencias entre ellos; en cambio, los modelos predictivos permiten estimar valores futuros o no conocidos de determinada variable. Desde el punto de vista de la gestión empresarial, los modelos obtenidos con esta técnica pueden contribuir a la toma de decisiones tácticas y estratégicas en una organización.

Por ello, esta investigación tiene como objetivo extraer conocimiento con la utilización de técnicas de minería de datos que permita describir la relación entre las partidas y elementos del presupuesto con el fin de apoyar la toma de decisiones en la elaboración del mismo en el ISMMM.

Durante el desarrollo de esta investigación fueron utilizadas las reglas de asociación para describir las relaciones entre los distintos elementos y partidas del presupuesto, los árboles de decisión para identificar cómo influyen en la ejecución anual del presupuesto el estado de ejecución de los distintos elementos y partidas. Finalmente, con el uso de la regresión lineal se obtuvo una función que estima el valor de la partida Total de gastos, lo que permitió establecer su correspondencia con otras partidas del presupuesto.

1. Antecedentes

El descubrimiento de conocimiento en las bases de datos es el proceso no trivial de identificación de patrones válidos, nuevos, potencialmente útiles y, finalmente, comprensibles en los datos, de acuerdo con Fayyad, Piatetsky-Shapiro & Smyth (1996, 40). En el centro de este proceso está la aplicación de métodos específicos de minería de datos para la extracción y descubrimiento de patrones, según Agrawal & Psaila (1995, 3), en una estructura comprensible para los humanos (Figura 1).

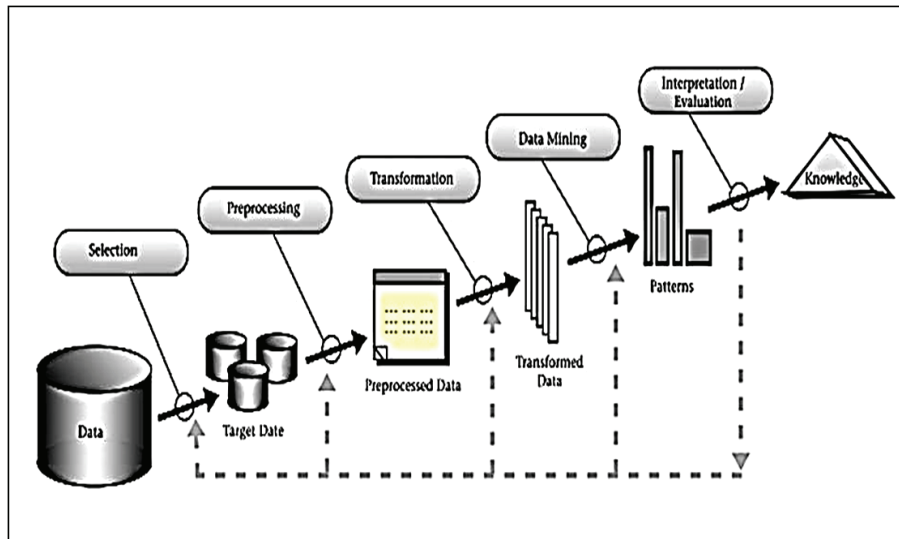


Figura 1. Pasos del proceso de descubrimiento de conocimiento en bases de datos (Fayyad, Piatetsky-Shapiro & Smyth, 1996, 41)

Existen varias metodologías y métodos para la realización de los proyectos de minería de datos, entre las principales metodologías utilizadas por los analistas están SEMMA (*Sample, Explore, Modify, Model, Assess*), propuesta por SAS Institute (1998), y CRISP-DM (*Cross-Industry Standard Process for Data Mining*) (Chapman et al., 2000, 6), que se encuentran actualmente entre los estándares metodológicos más difundidos para desarrollar un proceso de descubrimiento de conocimientos en bases de datos.

Estas metodologías estructuran el proyecto de minería de datos en fases interrelacionadas entre sí, convirtiendo el proceso en iterativo e interactivo. Asimismo, existe un número amplio de herramientas de apoyo al análisis de datos durante un proceso de descubrimiento de conocimientos en bases de datos. Generalmente, estas herramientas disponen de sus propios entornos gráficos y suelen permitir al usuario hacer múltiples tareas, pero siempre limitados a las especificaciones de cada aplicación. Son muchas las herramientas que sobresalen por su popularidad entre los desarrolladores de proyectos de minería de datos, entre las que se pueden mencionar *RapidMiner* (Mierswa et al, 2006), *SPSS Clementine* (SPSS Inc., 2008), *Excel* (Microsoft, 2007), *WEKA* (Hall et al., 2009), entre otros.

1.1 CRISP-DM

Fue propuesta inicialmente por un consorcio de empresas encabezadas por SPSS (*Statistical Product and Service Solutions*). En estos momentos constituye una de las metodologías más difundidas y utilizadas en proyectos de minería de datos (KDnuggets, 2007), pues propone un proceso estandarizado, no propietario y disponible libremente que permite adaptar la minería de datos y que esta sea incluida en la gestión empresarial o la investigación, como afirma Larose (2005, 5).

Esta metodología tiene seis fases fundamentales como guía durante el proceso de descubrimiento de conocimiento en bases de datos (Chapman et al., 2000):

- Comprensión del negocio: Se determinan los objetivos del negocio, se evalúa la situación y se elabora el plan del proyecto.
- Comprensión de los datos: Se recopilan los datos iniciales, se describen, se exploran y se verifica la calidad de los mismos.
- Preparación de los datos: Se seleccionan los datos, se limpian, construyen, integran y estructuran para ser minados.
- Modelado: Se seleccionan las técnicas de modelado, se genera el diseño del experimento y se construyen y evalúan los modelos.
- Evaluación: Se valoran los resultados, se revisa el proceso y se determinan las próximas acciones.

- Despliegue: Se traza la estrategia de empleo de los resultados, y se planifica el mantenimiento del proceso, además, se documentan las experiencias.

La metodología propone un orden para sus fases (Figura 2), aunque permite retrocesos entre varias de ellas, pues frecuentemente a lo largo del desarrollo de un proyecto, es necesario volver atrás en numerosas ocasiones, para reanализar los resultados obtenidos.

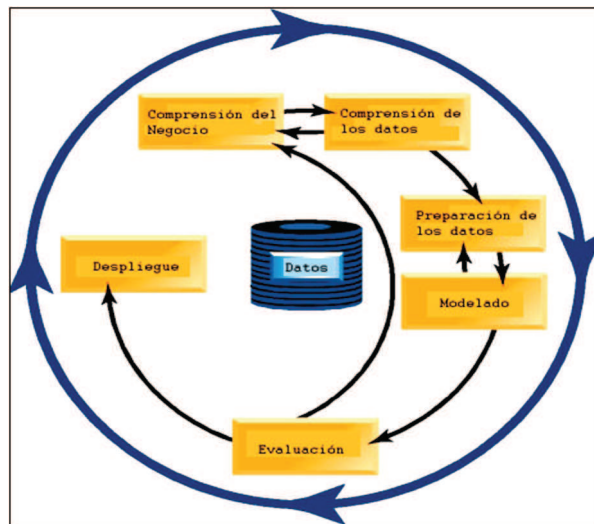


Figura 2. Metodología CRISP-DM (Chapman et al., 2000, 10)

1.2 WEKA (Waikato Environment for Knowledge Analysis)



Figura 3 Interfaz inicial del WEKA versión 3.4.15

El sistema, desarrollado en la Universidad de Waikato (Nueva Zelanda), es un software de libre distribución que implementa una colección de algoritmos de aprendizaje automático para tareas de minería de datos (Figura 3). Este sistema, señalan Witten & Frank (2000, 366), está escrito en Java, un lenguaje de programación orientado a objetos disponible para la mayoría de las plataformas computacionales. Además, establecen Witten & Frank (2000, 365), que existen diferentes niveles en los cuales WEKA puede ser utilizado. Ante todo, provee implementaciones de los algoritmos de aprendizaje y una amplia variedad de herramientas de transformación y preprocesado de datos.

2. Metodología

En la investigación se realizó el análisis de los datos de la gestión del presupuesto en el ISMMM utilizando técnicas de minería de datos. Para ello, inicialmente se efectuó un estudio comparativo entre las metodologías CRISP-DM (Chapman et al., 2000) y SEMMA (SAS, 1998) que considera, entre otros, los siguientes aspectos: estructuración del proceso, relación con la gestión empresarial y dependencia de las herramientas para el desarrollo de la minería de datos.

Aunque ambas metodologías están estructuradas en fases interrelacionadas, se decidió el empleo de la metodología CRISP-DM (Chapman et al., 2000) para guiar el proceso, pues la misma permite estructurar un proyecto de descubrimiento de conocimientos en bases de datos en fases interrelacionadas entre sí como un proceso iterativo e interactivo, que toma en cuenta los aspectos empresariales, tales como los objetivos del negocio. Además, no depende del uso de una herramienta específica y permite el contacto con los expertos en el área de conocimiento en la que se trabaja, los cuales brindan la información y el conocimiento necesario para la comprensión de los datos, la validación de los resultados obtenidos y la explotación de los modelos resultantes.

De igual manera, fue seleccionada la herramienta WEKA (Hall et al., 2009) para el procesamiento de los datos considerando que es de libre distribución, de código abierto y que puede ser utilizada con cualquier metodología de descubrimiento de conocimientos en bases de datos.

El proceso de extracción de conocimiento relativo a la gestión del presupuesto, se realizó teniendo en cuenta las fases que propone la metodología CRISP-DM.

2.1 Comprensión del negocio

El estudio realizado tuvo dos fines fundamentales: (1) determinar las relaciones que se establecen entre las partidas del presupuesto Total de gastos, Gastos de personal, Gastos de bienes y servicios, Otras transferencias corrientes y Gastos de capital y el ajuste al presupuesto planificado, y (2) determinar cómo influye el comportamiento de los epígrafes y partidas en la ejecución del presupuesto; con el fin de obtener un mayor conocimiento y dominio del comportamiento del presupuesto asignado, lo cual permitirá a la administración, si lo considera necesario, replantearse objetivos y metas.

Como criterios de éxito del cumplimiento de los objetivos fueron definidos los siguientes: la obtención de modelos que brinden una comprensión del comportamiento de las partidas, el aumento de los elementos que apoyan la toma de decisiones de los directivos, el informe sobre los problemas

encontrados en los datos, así como las recomendaciones para un mejor proceso de elaboración y control de la ejecución del presupuesto.

Todo ello permitió establecer los objetivos de la minería de datos: realizar una descripción inicial de los datos con el fin de lograr una mejor comprensión de los mismos y determinar los más significativos; obtener modelos que describan el comportamiento de las partidas del presupuesto, probar los modelos obtenidos.

Como colofón de esta fase fue elaborado el plan de proyecto, que incluyó la realización de entrevistas con los expertos, la revisión de la documentación relativa al presupuesto disponible en la Dirección de Economía, el estudio exploratorio de los datos, la extracción de patrones y la consulta de los resultados con los especialistas.

2.2 Comprensión de los datos

La información necesaria para el estudio fue recopilada de una única fuente, en la cual se registran los datos relacionados con el presupuesto. Los datos recopilados, contenidos en hojas de cálculo Excel, reflejan los valores asignados a las partidas y elementos del presupuesto y contiene datos históricos desde 2004 hasta el 2011. Asimismo, se obtuvo un resumen por partida en el que se refleja si el valor monetario asignado a la partida fue ejecutado en su totalidad (T), su ejecución estuvo sobregirada (SG), o fue incumplida (I), tomando como referencia lo que fue planificado.

De cada una de las partidas se obtuvo su valor máximo, valor mínimo, valor medio y la desviación estándar en el periodo estudiado.

2.3 Preparación de los datos

La preparación de datos está relacionada con las actividades necesarias para conformar el conjunto de datos final, en este estudio fueron eliminadas las partidas a las que no fue asignado presupuesto en el periodo 2004-2011, como es el caso de medicinas y materiales afines, donaciones corrientes, entre otras.

2.4 Modelado

Las técnicas de modelación seleccionadas estuvieron enfocadas a dar respuesta a los objetivos definidos, estas fueron:

Regresión, tarea predictiva de gran importancia, cuya meta es encontrar el valor numérico de la variable objetivo para objetos desconocidos (Hernández, Ramírez & Ferri, 2004, 26). Se propone la utilización del algoritmo de regresión lineal (Montgomery, Peek & Vining, 2007, 12), con el fin de obtener una función que permita predecir los valores de las partidas Total de gastos, Gastos de personal, Gastos de bienes y servicios, Otras transferencias corrientes y Gastos de capital.

Asociación, tarea descriptiva cuyo objetivo es identificar las relaciones no explícitas entre atributos nominales, se emplea frecuentemente para reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros (Hernández, Ramírez & Ferri, 2004, 26). Se utiliza el algoritmo Apriori (Agrawal, Imilienski & Swami, 1993, 209), con el fin de obtener reglas que describan las relaciones entre los distintos elementos y partidas del presupuesto, así como la relación de estos y su estado de ejecución.

Clasificación, considerada una de las tareas predictivas más importantes dentro de la minería de datos, tiene como objetivo organizar los datos en distintas clases a partir de crear un modelo basado en su distribución, de esta forma, puede ser clasificada con precisión nueva información (Hernández Ramírez & Ferri, 2004). Se propone la utilización del algoritmo J48, implementación que ofrece WEKA del algoritmo de obtención de árboles de decisión C4.5 (Quinlan, 1993, 17), con el fin de obtener cómo influyen en la ejecución anual del presupuesto el estado de ejecución de los distintos elementos y partidas, y de esta manera elaborar un sistema de reglas de ayuda a la toma de decisiones que permita realizar ajustes en la planificación en caso de ser necesario.

3. Resultados obtenidos

Se realizaron 33 experimentos, 11 por cada modelo a obtener. Para entrenar y probar los modelos se emplearon conjuntos distintos a fin de no sobrestimar su precisión, en este sentido se utilizó la validación cruzada de 10-pliegues (*10-folds cross validation*). En la figura 4 se muestran ejemplos de los resultados obtenidos en los experimentos realizados.

De la extracción de patrones utilizando la asociación, se obtuvo que en el periodo analizado se manifestaron regularidades como:

- Siempre que la partida Gastos de capital fue incumplida la partida Gastos de bienes y servicios fue ejecutada en su totalidad.
- Siempre que la partida Gastos de personal estuvo sobregirada la partida Gastos de bienes y servicios fue ejecutada en su totalidad.
- Siempre que la ejecución del presupuesto fue incumplida, la partida Gastos de personal fue incumplida.
- Siempre que la partida Gastos de bienes y servicios fue ejecutada en su totalidad, la partida Gastos de personal estuvo sobregirada en su ejecución y la partida Gastos de capital incumplida, se cumple con la ejecución del presupuesto.

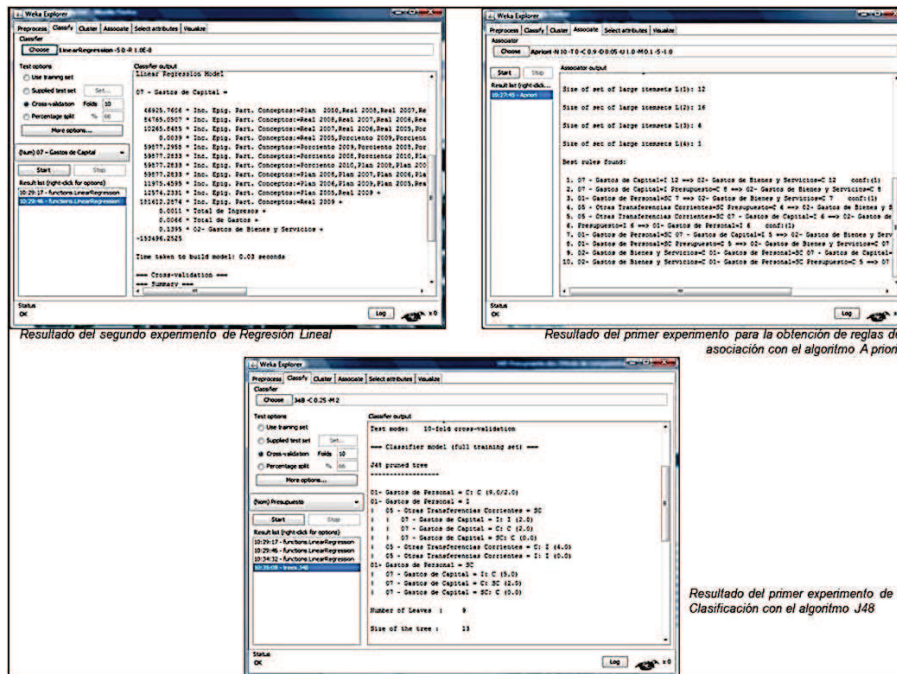


Figura 4. Algunos resultados obtenidos

La tarea de clasificación realizada, con el uso del algoritmo J48, permitió la obtención del árbol de decisión que se muestra en la figura 5.

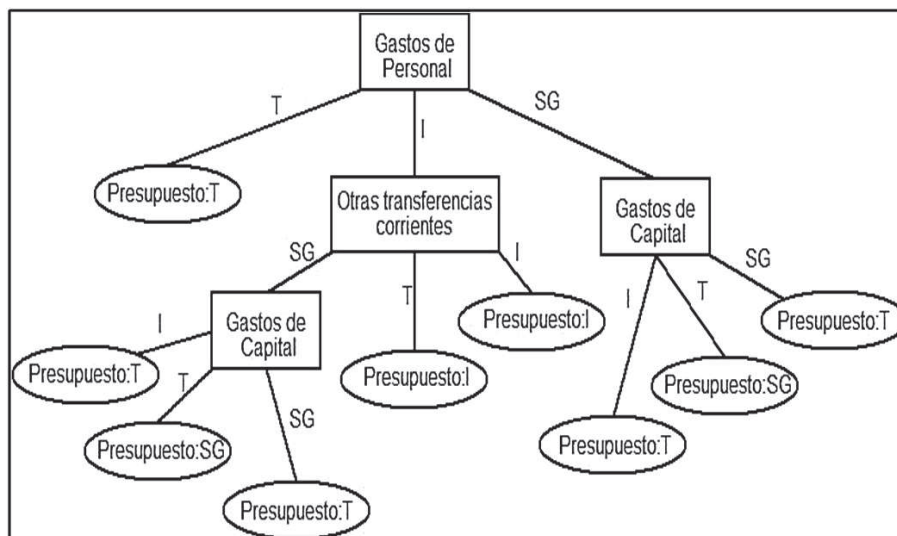


Figura 5. Árbol de decisión obtenido con el algoritmo J48.

Del árbol de decisión se obtienen reglas tales como: Si la ejecución de la partida Gastos de personal está sobregirada y la partida Gastos de capital se incumple entonces la ejecución del presupuesto se cumplirá en su totalidad. Esto permite recomendar a los especialistas de la Dirección de Economía: analizar las causas de la ejecución sobregirada de la partida Gastos de personal, y valorar realizar ajustes en la planificación incrementando el presupuesto asignado a Gastos de personal y reduciendo la asignación a la partida Gastos de capital.

Finalmente con la utilización de la regresión fueron obtenidos modelos como los siguientes:

- 1 $TG = 0.33TGC + 5.64GBS - 0.73GP + 0.75OTC - 0.53GC + 0.03$
- 2 $TG = 5.97GBS - 0.39GP + 1.08OTC - 0.53GC + 0.05$
- 3 $TGC = 0.17TG + 1.07GP + 0.82OTC + 0.09GC + 0.03$

Donde TG representa el total de gastos, TGC el total de gastos corrientes, GBS los gastos de bienes y servicios, GP los gastos de personal, OTC otras transferencias corrientes y GC los gastos de capital.

Tras el análisis de los modelos de regresión lineal obtenidos, fue seleccionado como más relevante el segundo modelo, pues permite predecir el total de gastos a partir de los valores estimados de gastos de bienes y servicios, gastos de personal, otras transferencias corrientes y gastos de capital, que dependen de la matrícula de pregrado y posgrado, la cantidad de profesores contratados a tiempo completo y a tiempo parcial, los materiales necesarios para la enseñanza, el financiamiento a la investigación, entre otros; lo que permite establecer el equilibrio que existe entre estas partidas.

Los modelos construidos tienen un coeficiente de correlación que sobrepasa el 0,7 y en el 87% de los modelos está por encima de 0,8. Los modelos obtenidos fueron utilizados en la planeación y análisis del presupuesto del año 2012 a la par del método usado tradicionalmente para corroborar su efectividad. Luego de la evaluación de los modelos y su presentación a los especialistas de la Dirección de Economía, fue elaborado un informe en el que se incluyen las recomendaciones de cómo pueden ser utilizados los patrones obtenidos.

La integración de los modelos obtenidos, la función de predicción del total de gastos en la fase de elaboración y las reglas de asociación y el árbol de decisión en la etapa de control de la ejecución, que incluye ajustes a la planificación de ser necesarios; desde el punto de vista analítico, sirve de apoyo a la toma de decisiones en la planificación y control de la ejecución del presupuesto, por lo que puede ser una he-

herramienta útil para los decisores y directivos del ISMMM en el proceso de gestión del presupuesto.

4. Conclusiones

El control presupuestario, es un proceso que permite evaluar la actuación y el rendimiento o resultado obtenido en cada entidad. El desarrollo de esta investigación siguiendo los pasos de la metodología CRISP-DM y la herramienta WEKA versión 3.4.15, permitió obtener modelos de regresión lineal, de reglas de asociación y de clasificación sobre la gestión de los presupuestos en el ISMMM durante el periodo 2004-2011.

El 80% de los modelos obtenidos en los experimentos realizados cumple con el criterio de tener un coeficiente de correlación superior a 0,8. El estudio de los elementos y partidas que conforman el presupuesto permitió comprender el comportamiento de los mismos.

La investigación realizada aportó nuevos conocimientos que mediante la integración de los modelos obtenidos con la utilización de la minería de datos, constituirá una herramienta útil para de apoyo a los directivos de la institución en la toma de decisiones relativas a la gestión del presupuesto en el ISMMM y en la estimación del presupuesto para años posteriores.

Referencias bibliográficas

- AGRAWAL, R. & PSAILA, G. (1995). Active Data Mining [online]. In: First International Conference on Knowledge Discovery and Data Mining, KDD-95 (20-21/08/1995), Montreal (Canada): Association for the Advancement of Artificial Intelligence, AAAI. FAYYAD, U.M. & UTHURUSAMY, R. (1995). Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95). AAAI Press. p. 3-8. ISBN: 0-929280-82-2 <<http://www.aaai.org/Papers/KDD/1995/KDD95-001.pdf>> [consult: 16/09/2013]
- AGRAWAL, R., IMIELINSKI, T. & SWAMI, A. (1993). Mining association rules between sets of items in large databases. In BUNEMAN, P. & JAJODIA, S. (ed.) (1993). In: SIGMOD '93 (26-28/05/1993), Washington D.C. (USA): ACM. Proceedings of the ACM SIGMOD International Conference on Management of Data, New York (NY, USA): ACM Press, p. 207-216. ISBN: 0-89791-592-5.
- BURBANO RUIZ, J. E. (2005). Presupuestos. Enfoque de gestión, planeación y control de recursos. 3 ed. Santa Fe de Bogotá (Colombia): McGraw-Hill. 405 p. ISBN: 978-9584103598.
- BURBANO RUIZ, J. E. & ORTIZ GÓMEZ, A. (1995). Presupuestos: Enfoque moderno de planeación y control de recursos. 2 ed. Santa Fe de Bogotá (Colombia): McGraw-Hill. 376 p. ISBN: 9789586004473.
- CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHAHAZA, T.; REINARTZ, T.; SHEARER, C. & WIRTH, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide [online]. Chicago (IL, USA): SPSS Inc. 73 p. <<ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>> [consult: 16/09/2013]

- FAYYAD, U.; PIATETSKY-SHAPIRO, G. & SMYTH, P. (1996). From Data Mining to Knowledge Discovery in Databases [online]. In: AI Magazine, Vol.17, No.3 (fall). Palo Alto (CA, USA) The Association for the Advancement of Artificial Intelligence (AAAI), p. 37-55. <<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>> [consult: 16/09/2013]
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P. & WITTEN, I. H. (2009). The WEKA Data Mining Software: An Update [online]. In: SIGKDD Explorations, Vol. 11, No. 1 (jun). New York (NY, USA): ACM. p. 10-18. <<http://dl.acm.org/citation.cfm?id=1656274.1656278>><doi: 10.1145/1656274.1656278>[consult: 16/09/2013]
- HERNÁNDEZ ORALLO, J.; RAMÍREZ QUINTANA, M.J. & FERRI RAMÍREZ, C. (2004). *Introducción a la minería de datos*. Madrid (España): Pearson Educación. 680 p. ISBN: 84 205 4091 9
- KDNUGGETS (2007). 2007 KDNuggets Poll on Data Mining Methodology [on line]. Boston (MA, USA): KDNuggets.com. <http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm> [consult: 23/11/2013]
- KIRKOS, E.; SPATHIS, C. & MANOLOPOULOS, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements [online]. In: Expert Systems with Applications. 2007. Vol. 32, No. 4 (may). New York (NY, USA): Elsevier. p. 995–1003. ISSN: 0957-4174<<http://delab.csd.auth.gr/papers/ESWA07ksm.pdf>> [consult: 16/09/2013]
- LAROSE, D. T. (2005). *Discovering Knowledge in data: An introduction to data mining*. Hoboken (NJ, USA): John Wiley & Sons (Wiley-Interscience Publication). 222 p. ISBN: 0-471-66657-2
- MICROSOFT (2007). Microsoft Excel 2007 [en línea]. México D.F. (México): Microsoft Corporation Mexico. <<http://office.microsoft.com/es-mx/excel/>> [consulta: 16/09/2013]
- MIERSWA, I.; WURST, M.; KLINKENBERG, R.; SCHOLZ, M. & EULER, T. (2006). Yale: rapid prototyping for complex data mining tasks. In: Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'06 (20-23/08/2006), Philadelphia (USA): ACM. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06. p. 935-940. New York (NY, USA): ACM. ISBN: 1-59593-339-5
- MONTGOMERY, D. C.; PEEK, E. A. & VINING, G. G. (2006). *Introduction to linear regression analysis* (Wiley Series in Probability and Statistics). 4 ed. Hoboken (NJ, USA): John Wiley & Sons (Wiley-Interscience Publication). 640 p. ISBN: 978-0471754954
- NELSON, K. (2006). Budgeting. In: HELMS, M. M. (ed.) (2006). *Encyclopedia of Management*. 5 ed. Farmington Hills (MI, USA): Thomson Gale. p. 49-50. ISBN: 0-7876-6556-8
- QUINLAN, J. R. (1993). *C4.5. Programs for Machine Learning*. San Francisco (CA, USA): Morgan Kaufmann. 302 p. ISBN: 1-55860-238-0
- SAS INSTITUTE INC. (1998). *SAS Institute White Paper, from data to business advantage: Data Mining, The SEMMA Methodology and the SAS® System*. Cary (NC, USA): SAS Institute Inc.
- SPSS INC. (2008). SPSS Clementine 12.0 [online]. Chicago (IL, USA): SPSS Inc. <<http://www.tgc.com/dsstar/03/1014/106790.html>> [consult: 16/09/2013].
- SWEENEY, A. H. W. & RACHLIN, R. (1984). *Manual de presupuestos*. México DF (México): McGraw-Hill. 884p. ISBN: 978-9684515529
- WITTEN, I. H. & FRANK, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 2 ed. Collection: The Morgan Kaufmann Series in Data Management Systems. Waltham (MA, USA): Morgan Kaufmann Publishers. 371 p. ISBN: 978-1558605527

Bibliografía complementaria

- SULLIVAN, A. & SHEFFRIN, S. M. (2005). *Economics: Principles in action*. Upper Saddle River (NJ, USA): Pearson Education. 592 p. ISBN: 978-0131334830.