



**La detección del fraude contable utilizando técnicas de Minería de datos**

**Alexander Báez Hernández<sup>1</sup>**

**1 Universidad Central del Ecuador, [albaez@uce.edu.ec](mailto:albaez@uce.edu.ec)**

**RESUMEN**

Se realizó una revisión bibliográfica en ScienceDirect para los términos “Data Mining” AND “fraud” lo que permitió seleccionar un total de 86 artículos para el periodo 2000 al 2015. Esta revisión comprobó que aplicación de las técnicas de la Minería de Datos para la detección de los distintos tipos de fraude contable es una tendencia de investigación actual que puede ser de gran utilidad. La Minería de Datos permite realizar los procesos de clasificación, detección de valores anómalos., agrupamiento, predicción y regresión, todos necesarios para poder tratar los datos contables.

Se aplicó RapidMiner para poder ejemplificar la detección de valores anómalos y el agrupamiento por clústeres de una data de prueba seleccionada de 1000 datos. Las posibilidades de visualización existentes permiten a la vez una fácil detección, tanto de los posibles valores anómalos como del posible agrupamiento de los datos.

La investigación realizada sugiere la combinación de las posibilidades de la Minería de Datos en el análisis de un mayor volumen de datos y a la vez la introducción de estas técnicas en la docencia actual de la Contabilidad.

**Palabras claves:** fraude contable, minería de datos, RapidMiner, detección valores anómalos, agrupamiento en clúster



**Detection of accounting fraud utilizing Data Mining Techniques**

**ABSTRACT**

A literature review on ScienceDirect for the terms "Data Mining" AND "fraud" allowed selecting 86 papers for the period 2000 to 2015. This review allowed finding that the application of data mining techniques is useful to detect the different types of accounting frauds. In addition, is also a current research topic. Data mining allows developing the processes of: classification, detection of outliers, clustering, forecasting and regression, all necessary to treat the accounting data.

RapidMiner was applied to illustrate the detection of outliers and clustering of the selected test data. The existing visualization possibilities allow also and easy detection of outliers and the clustering of data.

The research suggests a combination of the possibilities of data mining in the analysis of a greater volume of data and the introduction of these techniques in the current teaching of Accounting.

**Keywords:** Accounting fraud, data mining, RapidMiner, outliers detection, clustering



## **1. INTRODUCCIÓN**

La Minería de Datos es el proceso analítico de exploración de los datos en la búsqueda de patrones consistentes, o relaciones sistemáticas entre las variables y para poder validar y aplicar los patrones encontrados a un nuevo conjunto de datos (Software, 2015).

Las técnicas de la Minería de Datos han sido aplicadas en los más variados entornos desde la educación (Zengin, Esgi, Erginer, & Aksoy, 2011), las relaciones con clientes (Ngai, Xiu, & Chau, 2009), la inteligencia de negocios (Chen, Chiang, & Storey, 2012) y los estudios bibliométricos (González-Alonso, Fabian-Pazmiño, & Pérez-González, 2015).

El estudio del fraude y su investigación es un campo natural para la aplicación de las técnicas de la Minería de Datos (Phua, Lee, Smith, & Gayler, 2010). La detección de anomalías en los procesos de creación, almacenamiento, registro y transmisión de un conjunto de datos de un punto a otro es otro escenario para la aplicación de las técnicas de Minería de Datos (Agrawal & Agrawal, 2015). El escenario económico actual en que se presentan distintos tipos de fraudes en las finanzas y la contabilidad ha también propiciado la aplicación de las técnicas de la Minería de Datos para la detección de las posibles fallas en los sistemas contables (Sharma & Panigrahi, 2013)

Dentro de esta problemática este reporte tuvo como objetivo realizar una revisión del estado del arte sobre el tema de la aplicación de las técnicas de la Minería de Datos en la detección del fraude contable. A partir de esta revisión se procedió a realizar un estudio exploratorio empleando un paquete de Minería de Datos (RapidMiner, 2014) para el análisis de datos primarios de Contabilidad, para la detección de valores atípicos (outliers) y para el agrupamiento en clústeres de estos datos.



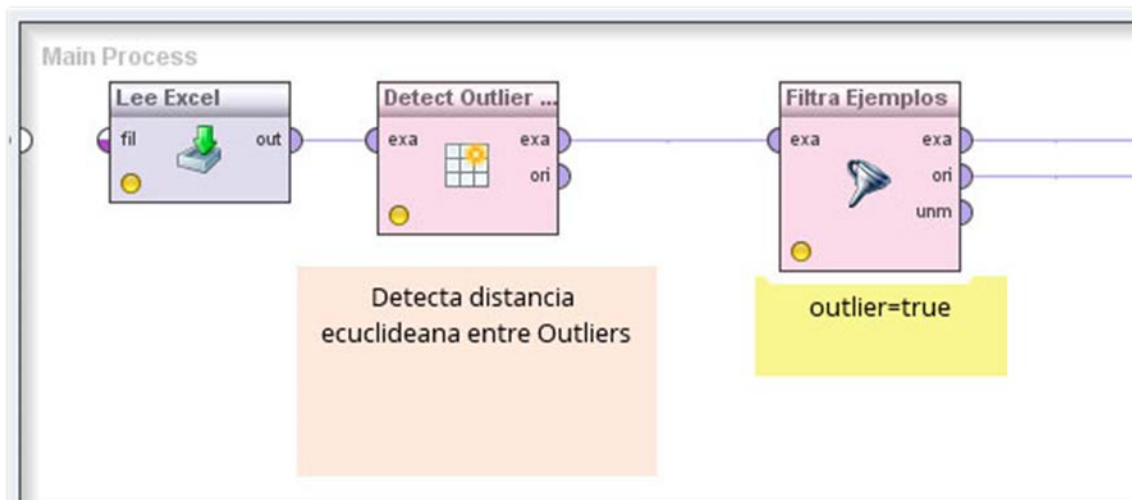
## 2. METODOS

Dado lo relevante del tema a explorar se decidió realizar la revisión para los términos: “Data Mining” AND “fraud” empleando ScienceDirect (ScienceDirect, 2015) ello permitió seleccionar un total de 86 artículos para el periodo 2000 al 2015.

Para la aplicación de las técnicas de Minería de Datos, se decidió comprobar dos de ellas: detección de valores atípicos y agrupamiento en clústeres, para ello se empleó RapidMiner por la facilidad que ofrece este paquete de poder utilizar objetos ya programados.

El esquema de procesos para la detección de valores anómalos se muestra en la Figura 1:

Figura 1: Detección de valores anómalos empleando RapidMiner



Como se observa consta de tres módulos

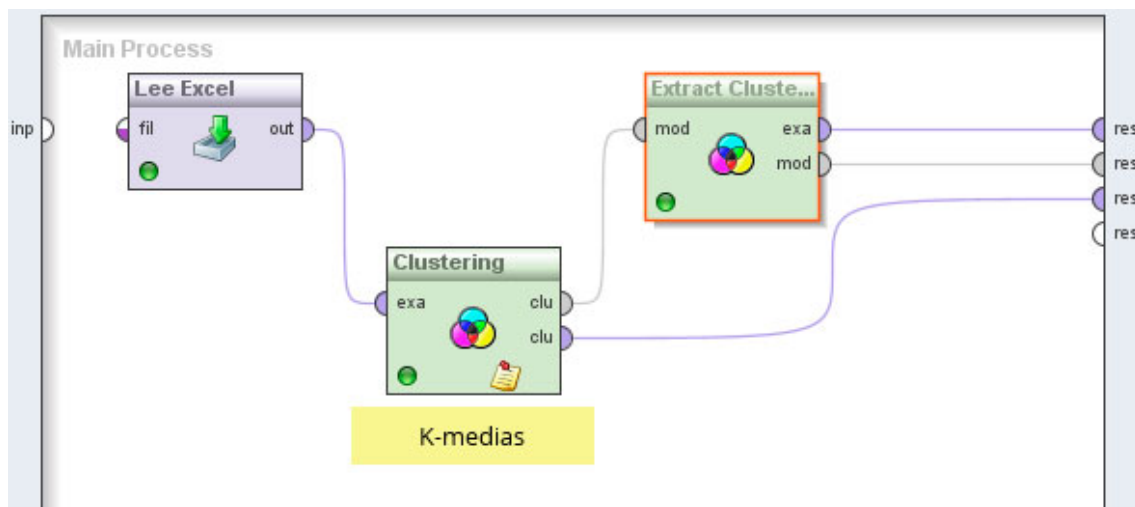
- 1) Lectura de los Datos en Excel. Se tomó un conjunto de 1000 datos, para los que se tenía: Identificación del Documento Primario, Fecha de la transacción, Monto de esta.
- 2) Detección de valores anómalos. Este módulo ya está implementado en RapidMiner y se seleccionó la detección de distancia euclidiana



- 3) Filtrado. Se realizó para el valor outlier= true, lo que permite obtener la Tabla de valores anómalos

Para el agrupamiento en clústeres se utilizó el esquema de procesos siguientes:

Figura 2. Agrupamiento en clústeres empleando RapidMiner



Los tres módulos son:

- 1) Lectura de los Datos en Excel. Se tomó el mismo conjunto de 1000 datos
- 2) Agrupamiento en clústeres. Se empleó el algoritmo de k.medias
- 3) Extracción del prototipo de clúster

### **3. RESULTADOS**

#### **3.1 Revisión de la literatura**

La revisión realizada comprobó que la aplicación de las técnicas de la Minería de datos es una tendencia de investigación actual que puede ser de gran utilidad para la detección de los distintos tipos de fraude contable.

En relación con los métodos estadísticos de detección de fraude (Bolton & Hand, 2002) la Minería de Datos aporta (Zhou, 2003) además de la perspectiva estadística, la del manejo de grandes volúmenes de datos y la posibilidad de aprendizaje a partir de la misma data unido a nuevas formas de visualización de la misma.



(Sharma & Panigrahi, 2013) Presentan un esquema conceptual de gran utilidad para la aplicación de la Minería de Datos en la detección del fraude financiero. Ver Figura 3.

Figura 3. Marco Conceptual para la aplicación de la Minería de Datos en el análisis del fraude financiero



Basado en (Sharma & Panigrahi, 2013)

Como se observa los paquetes actuales de Minería de Datos pueden ofrecer la visualización conjunta de la técnica aplicada.

Un aspecto importante señalado en la literatura (Phua et al., 2010) es las posibilidades del aprendizaje supervisado a partir de una data y a la vez la necesaria interrelación de la Minería de Datos con otras disciplinas e incluso entre las distintas aplicaciones de esta.

Las aplicaciones reportadas recientemente comprenden el fraude en tarjetas de crédito (Zareapoor & Shamsolmoali, 2015), en los sistemas de salud (Thornton, Brinkhuis, Amrit, & Aly, 2015) así(Trigo, Belfo, & Estébanez, 2014) como los retos que implica la detección en tiempo real en los sistemas contables (Trigo et al., 2014), autores como



(Belfo & Trigo, 2013) marcan tanto el reto, como el hecho de que aún es deficiente en la profesión contable el manejo de estas nuevas tecnologías

### **3.2 Ejemplos de aplicación de técnicas de Minería de Datos**

A partir de la revisión realizada y tomando como base el esquema presentado en la Figura 3 se procedió a aplicar dos de ellas para el análisis de 1000 datos.

#### **a) Detección de valores anómalos**

A partir del esquema que se presentó en la Figura 1 se procedió a detectar valores anómalos al conjunto de 1000 datos tomados para la prueba.

Los valores anómalos se presentan en la Tabla 1

Tabla 1. Valores anómalos obtenidos para los 1000 Documentos primarios procesados

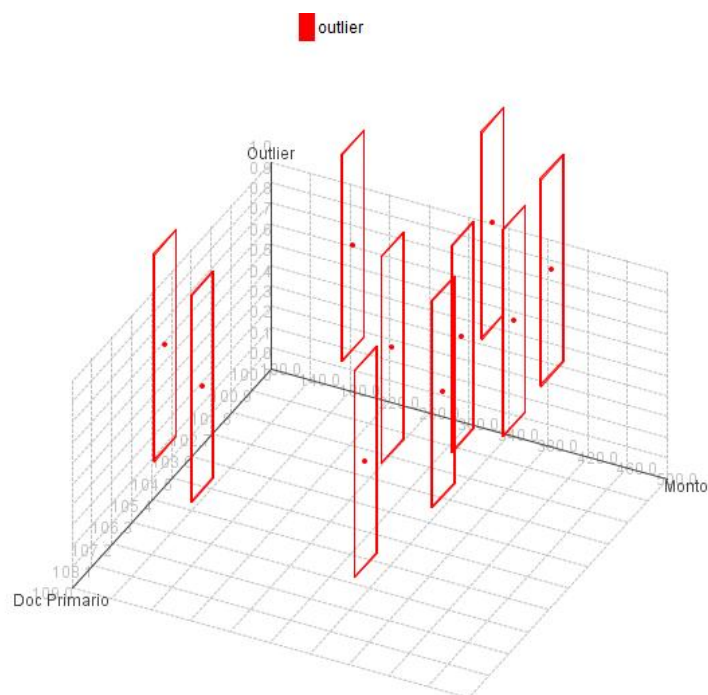
Outlier	Doc. Primario	Fecha	Monto
true	100	Mon Jun 01	324,15
true	101	Mon Jun 01	406,46
true	102	Mon Jun 01	227,76
true	103	Mon Jun 01	412,54
true	104	Mon Jun 01	382,97
true	105	Mon Jun 01	334,65
true	106	Mon Jun 01	408,11
true	107	Mon Jun 01	149,27
true	108	Mon Jun 01	209,40
true	109	Mon Jun 01	396,53

Las posibilidades de visualización se observan en la Figura 4, en que se han representado los valores anómalos y su desviación estándar en el espacio Monto-Documento Primario-Outlier.



Esta visualización evidencia una de las ventajas de la Minería de Datos que permite representar y observar un conjunto de datos que puede ser relativamente grande y permitir al contador concentrarse de forma muy rápida en las anomalías.

Figura 4. Valores anómalos y su desviación estándar



#### **b) Agrupamiento en clústeres**

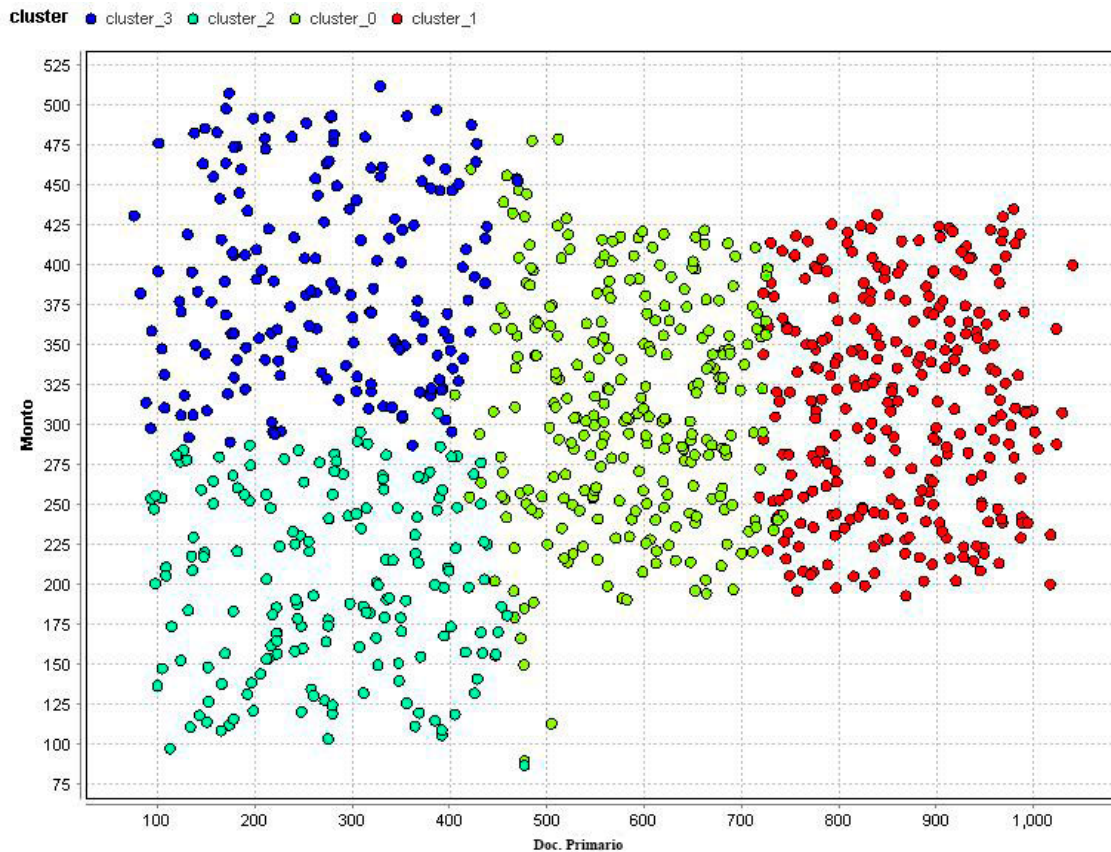
A partir de los procesos que se detallaron en la Figura 2 se procedió a agrupar en cuatro clústeres todos los datos procesados. Los resultados se presentan en la Figura 5 y





presentan los datos desde otra perspectiva que permiten analizar las causas y motivación de la separación en los clústeres visualizados y diferenciados por colores

Figura 5. Agrupamiento en clústeres de todos los datos analizados



Como se observa los dos ejemplos analizados ejemplifican la potencialidad de las técnicas de Minería de Datos tanto en la factibilidad de procesamiento de los datos como en la visualización de estos.

#### 4. CONCLUSIONES

La revisión de la literatura permitió comprobar que la aplicación de las técnicas de la Minería de datos puede ser de gran utilidad para la detección de los distintos tipos de fraude contable y una línea de investigación de gran perspectiva. La Minería de Datos permite la clasificación, detección de valores anómalos., el agrupamiento, la predicción y la regresión necesarios para poder tratar los datos contables.



El RapidMiner resultó de fácil aplicación para poder ejemplificar la detección de valores anómalos y el agrupamiento por clústeres de una data de prueba seleccionada de 1000 datos. Las posibilidades de visualización existentes permiten a la vez una fácil detección tanto de los posibles valores anómalos como del posible agrupamiento de los datos.

La investigación realizada sugiere la combinación de las posibilidades de la Minería de Datos en el análisis de un mayor volumen de datos y a la vez la introducción de estas técnicas en la docencia actual de la Contabilidad

## **5. REFERENCIAS BIBLIOGRÁFICAS**

- Agrawal, S., & Agrawal, J. (2015). Survey on Anomaly Detection using Data Mining Techniques. *Procedia Computer Science*, 60, 708-713. doi: <http://dx.doi.org/10.1016/j.procs.2015.08.220>
- Belfo, F., & Trigo, A. (2013). Accounting Information Systems: Tradition and Future Directions. *Procedia Technology*, 9, 536-546. doi: <http://dx.doi.org/10.1016/j.protcy.2013.12.060>
- Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science*, 235-249.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, 36(4), 1165-1188.
- González-Alonso, J. A., Fabian-Pazmiño, J., & Pérez-González, Y. (2015). Posicionamiento en Google Académico y en la WEB de la Revista Ciencias Holguín. *Ciencias Holguín*, 21(2), 44-54.
- Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592-2602.
- Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.
- RapidMiner. (2014). Products. RapidMiner Studio. Retrieved 8 Enero, 2014, from <http://rapidminer.com/products/rapidminer-studio/>
- ScienceDirect. (2015). ScienceDirect. Retrieved Junio, 2015, from <http://www.sciencedirect.com/>



- Sharma, A., & Panigrahi, P. K. (2013). A review of financial accounting fraud detection based on data mining techniques. *arXiv preprint arXiv:1309.3944*.
- Software, D. (2015). Data Mining Techniques. *Statistics – Textbook*. Retrieved Agosto 15, 2015, from <http://documents.software.dell.com/Statistics/Textbook/Data-Mining-Techniques#mining>
- Thornton, D., Brinkhuis, M., Amrit, C., & Aly, R. (2015). Categorizing and Describing the Types of Fraud in Healthcare. *Procedia Computer Science*, 64, 713-720. doi: <http://dx.doi.org/10.1016/j.procs.2015.08.594>
- Trigo, A., Belfo, F., & Estébanez, R. P. (2014). Accounting Information Systems: The Challenge of the Real-time Reporting. *Procedia Technology*, 16, 118-127. doi: <http://dx.doi.org/10.1016/j.protcy.2014.10.075>
- Zareapoor, M., & Shamsolmoali, P. (2015). Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier. *Procedia Computer Science*, 48, 679-685. doi: <http://dx.doi.org/10.1016/j.procs.2015.04.201>
- Zengin, K., Esgi, N., Erginer, E., & Aksoy, M. E. (2011). A sample study on applying data mining research techniques in educational science: Developing a more meaning of data. *Procedia - Social and Behavioral Sciences*, 15, 4028-4032. doi: <http://dx.doi.org/10.1016/j.sbspro.2011.04.408>
- Zhou, Z.-H. (2003). Three perspectives of data mining. *Artificial Intelligence*, 143(1), 139-146. doi: [http://dx.doi.org/10.1016/S0004-3702\(02\)00357-0](http://dx.doi.org/10.1016/S0004-3702(02)00357-0)