

Minería de Datos para la prevención del lavado de dinero

Alarcón Peralta L. A.¹, Ramírez Meléndez A.², Ávila Camacho F.J.³



requerimientos pueden ser remediados por medio de técnicas de minería de datos que, trasladadas a los sistemas de detección, jugarán un papel muy importante, pues a través del análisis y conocimiento que producen, se busca prevenir y detectar la mayor parte de las actividades ilícitas y, por consecuencia, contener sus impactos. Ante un creciente volumen de información por centralizar y estudiar, y en aras de desempeñar una labor más efectiva, se hace necesario emplear técnicas de análisis de datos más complejas y dinámicas a las tradicionales, para lograr, finalmente, la extracción de conocimiento no implícito.

Palabras Clave: Algoritmos, Minería de Datos, Árboles de Decisión, Redes Neuronales, Rapidminer.

Abstract

Financial globalization, as well as tighter regulation in the economic market, have caused governments to pressure financial institutions to move forward in developing systems for the detection and prevention of money laundering that allow them to be one step ahead of violators, improving their detection and updating how unusual transactions are warned as anti-money laundering laws are being updated; these requirements can be remedied by means of data mining techniques that, transferred to detection systems, will play a very important role, because through the analysis and knowledge they produce, it seeks to prevent and detect most of the illegal activities and, consequently, contain their impacts. In the face of a growing volume of information to be centralized and studied, and in order to carry out a more effective work, it is necessary to employ more complex and dynamic data analysis techniques to traditional ones, in order to finally achieve the extraction of unconced knowledge.

Keywords

Algorithms, Data Mining, Decision Trees, Neural Networks, Rapidminer.

Introducción

El manejo de los activos financieros es parte fundamental de cualquier institución gubernamental o privada, así como de la tecnología para la realización de sus actividades en forma rápida y eficiente. Actualmente, las leyes y disposiciones para la prevención de lavado de dinero son un tema de vital importancia en nuestros días, ya que las entidades gubernamentales y privadas tienen la obligación legal de reportar y luchar contra este fenómeno. El lavado de dinero se desplaza y adapta continuamente en el mercado financiero, impidiendo que las reglas establecidas puedan detectar todas las transacciones ilícitas.

El hecho de que las leyes en contra de estas actividades sean más sólidas y deban actualizarse con rapidez y dinamismo, ha implicado que se busquen maneras más eficaces de detectar transacciones ilícitas de formas expedita y confiable. Es importante que se apliquen nuevas tecnologías, como la minería de datos y la inteligencia artificial, para detectar y evitar transacciones no deseadas, logrando ofrecer un panorama amplio que permita no solamente detectar y prevenir, sino persuadir posibles fraudes y enriquecimientos ilícitos, que muchas veces y a pesar de los esfuerzos, pasan inadvertidos en los sistemas actuales. Utilizando redes neuronales y árboles de decisión, se buscará un modelo de clasificación que divida un conjunto de análisis, buscando el mayor grado de pureza entre los datos resultantes.

1. Prevención de lavado de dinero

El término lavado de dinero, se refiere a las actividades financieras que se realizan

Resumen

La globalización financiera, así como una regulación más estricta en el mercado económico, han ocasionado que los gobiernos presionen a las entidades financieras a avanzar en el desarrollo de sistemas para la detección y prevención de lavado de dinero que les permitan estar un paso adelante de los infractores, mejorando su detección y actualizando el modo de advertir transacciones inusuales conforme se van actualizando las leyes en contra del lavado de dinero; estos

Acerca de los autores...

^{1,2,3} Div. De Ingeniería en Sistemas Computacionales, Tecnológico de Estudios Superiores de Ecatepec.

con el fin de ocultar el origen verdadero de los fondos recibidos, que muchas veces provienen de actividades ilegales, y su objetivo es darle a ese dinero ilegal la apariencia de tener su fuente en una actividad legalmente constituida, y una vez efectuado este proceso, estar disponible para su utilización por el crimen organizado.

1.1 La detección de fraudes y sus problemas

El problema en la detección de fraudes radica en el análisis de perfiles que permitan conocer el comportamiento de un cliente, con el fin de detectar anomalías. En Customer Resource Management (CRM), el análisis de la información de un cliente implica una cadena de datos, como se muestra a continuación.

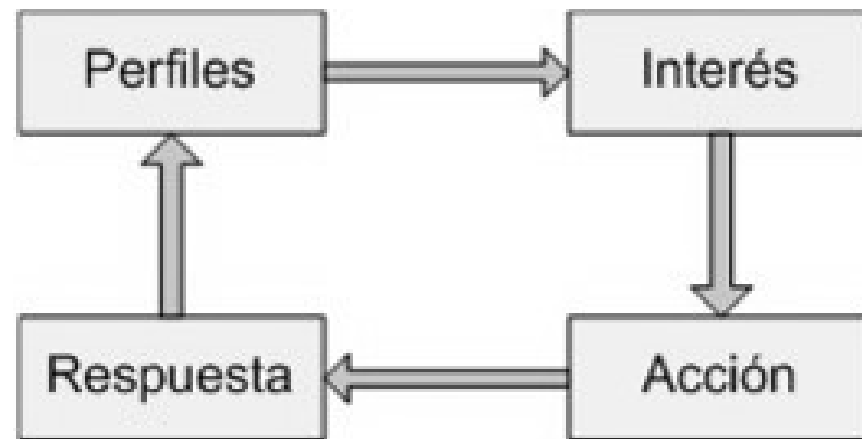


Figura 1

Cadena de datos CRM para el análisis de clientes.

1.2 Técnicas para la detección de fraudes

Actualmente las herramientas para la detección de un fraude se clasifican en:

- Técnicas tradicionales.
- Técnicas de minería de datos.

En la siguiente tabla, se mencionará el objetivo y la finalidad de las técnicas tradicionales y de minería de datos utilizadas en la detección de registros inusuales para la prevención del lavado de dinero.



Tabla 1
TÉCNICAS PARA LA DETECCIÓN DE FRAUDES

Técnicas Tradicionales y/o de Minería de Datos	Objetivo	Finalidad
- Análisis de anomalías.	Descubrir datos inusuales.	- Detectar registros con valores anormales. - Detectar varias ocurrencias de valores. - Detectar relaciones entre registros.
- Análisis de Clúster y anomalías. - Análisis de relaciones de asociación.	Reconocer relaciones inexplicables	- Determinar perfiles. - Determinar registros con referencias anormales. - Detectar relaciones indirectas entre registros. - Detectar registros con combinaciones de valores anormales.
- Modelos predictivos.	Conocer las características del fraude.	- Encontrar criterios y reglas. - Calificación de transacciones sospechosas.

2. Inteligencia artificial

La inteligencia artificial comenzó como el resultado de la investigación en psicología cognitiva y lógica matemática. Se enfoca en la explicación del trabajo mental y la construcción de algoritmos de solución a problemas de propósito general, favoreciendo la abstracción y la generalidad.

2.1 Redes neuronales

En las dos últimas décadas se ha visto un gran número de investigaciones y aplicaciones de las Redes Neuronales para una variedad de problemas de clasificación del mundo real. **Algunos estudios empíricos han señalado que hay algunos dominios de problemas para los cuales las redes neuronales ofrecen una precisión superior de clasificación en comparación a otros algoritmos de aprendizaje.**

2.2 Árboles de decisión

Un árbol de decisión es un modelo de clasificación que divide un conjunto de análisis, buscando el mayor grado de pureza entre los grupos resultantes. En un árbol hay un nodo inicial, denominado raíz, que contiene la totalidad de la información. Este grupo se divide en dos o más grupos, llamados internos, si continúa subdividiéndose, o terminales u hojas, si no enfrentan más segmentación.

3. Modelo Knowledge Discovery in Database KDD

A continuación, se expone el modelo que será empleado por la unidad financiera en la generación de conocimiento para detectar operaciones de lavado de dinero.

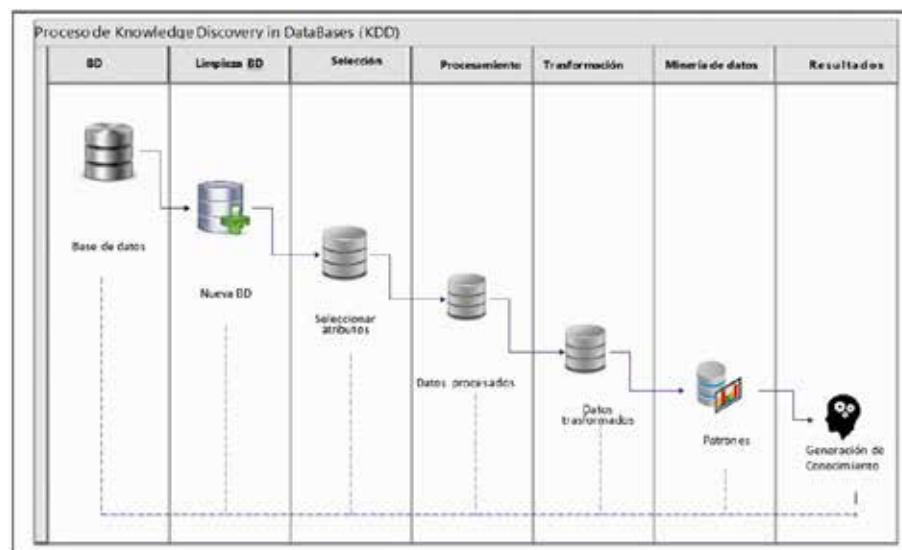


Figura 2

Cadena de datos CRM para el análisis de clientes

El proceso más importante es el de limpieza de los datos y definición de las variables. Si los datos no son correctos, el modelo creado no servirá. De igual modo, la validez de los patrones descubiertos depende de cómo se apliquen al mundo real o a las circunstancias. La eficacia de este enfoque se ve limitada por la estructura del software que utiliza. La minería de datos emplea una aproximación de descubrimiento para examinar simultáneamente varias relaciones en bases de datos multidimensionales, identificando aquellas que se presentan con frecuencia. Luego se analizan los resultados de los patrones obtenidos en la fase de minería de datos, mediante técnicas de visualización y de representación, con el fin de generar conocimiento que aporte mayor valor a los datos. En esta fase, se evalúan los resultados con los expertos y, si es necesario, se retorna a las fases anteriores para una nueva iteración. **Las actividades mencionadas anteriormente se resumirán.**

4. Aplicación de modelo KDD en la empresa

Selección de los datos: se realiza la recolección y preparación de los mismos desde una base de datos empresarial; en este proceso, se comprende la problemática asociada a dicha base y se establecen los objetivos.

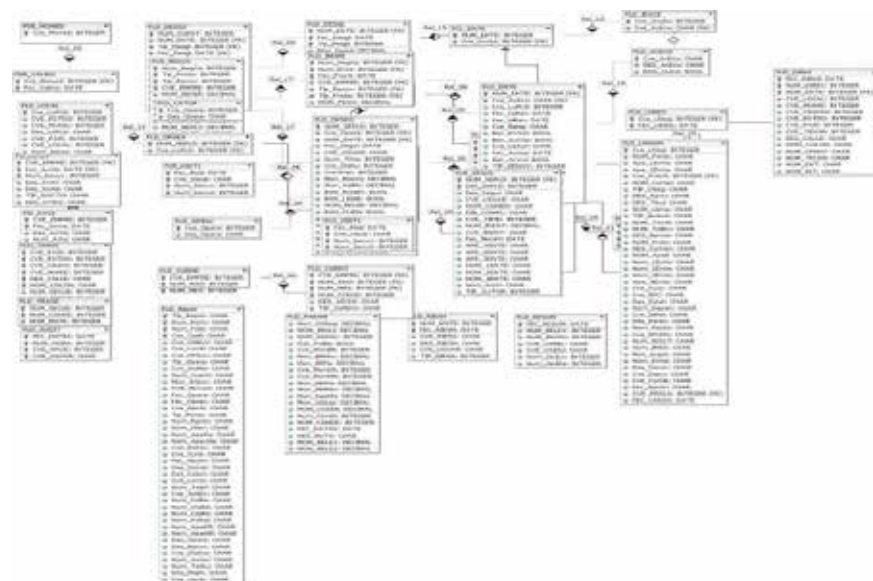


Figura 3

Selección de datos de una base de datos.

Pre procesamiento de datos: consta de la *integración de datos* analizando si la BD requiere incluir o integrar información.

Reconocimiento y limpieza, donde se depura el conjunto de datos, eliminando ruidos e inconsistencias.

La imagen muestra una interfaz de usuario con una tabla de datos. Las columnas incluyen 'Ord_id', 'No. Cliente', 'CURP', 'IFE', 'A. Paterno', 'A. Materno', 'Fecha Recuento', 'Estado Recuento', 'Sexo' y 'Tratamiento'. Las filas muestran datos de clientes con valores como '144.252', 'RELD77...', 'RYLPDL...', 'REYES', 'LOPEZ', 'DILMA'.

Selección de características: consiste en una exploración y limpieza, aplicando técnicas de análisis exploratorio de datos.

Figura 4

Integración de datos, reconocimiento y limpieza.

Transformación: estandarización o normalización de información.

Reducción de datos: selección extracción de atributos, realizando un proceso de identificación y selección de variables.

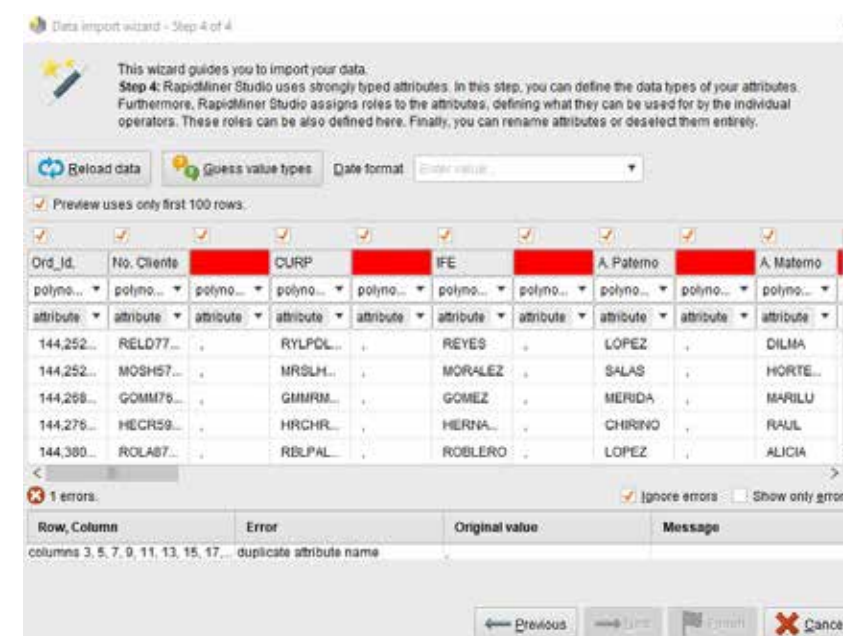


Figura 5

Exploración y limpieza de datos, reducción de datos, transformación.

Minería de datos: proceso no trivial de identificación, potencial de patrones comprensibles, donde se encuentran ocultos en los datos que facilitarán la toma de decisiones, empleando técnicas de aprendizaje supervisado y no supervisado.

Name	Type	Missing	Statistics	Filter (30 / 30 attributes)
Ord_id	Integer	0	Min: 144, Max: 144, Average: 144	
No. Cliente	Integer	0	Min: 163205, Max: 400204, Average: 290211.130	
CURP	Polynomial	0	Labels: VERE51011MCSLDL0, AEBK910815MCSNLR0, AEBK910815MCSNLR0 (1), AEBV930219MK	
IFE	Polynomial	0	Labels: VRHRCR75071707M70, AGBSGN52121607M00, AGBSGN52121607M00 (1), ANBLKR910815	
A. Paterno	Polynomial	0	Labels: VELAZQUEZ (1), MENDEZ (5), MENDEZ (5), ESPINOSA (3), ... [39 more]	
A. Materno	Polynomial	0	Labels: URBANA (1), LOPEZ (5), LOPEZ (5), GARCIA (3), ... [36 more]	
Nombre (s)	Polynomial	0	Labels: ZULEMA (1), ALICIA (2), ALICIA (2), GUADALUPE (2), ... [49 more]	

Figura 6

Identificación, clasificación, agrupación, asociación, predicción.

Interpretación y resultados: se analizarán los patrones obtenidos, mediante técnicas de visualización y de representación, con el fin de generar conocimiento.



Tree

Personal Trabajando > 0.500: 3,360.00 {1,000.00=0, 1,500.00=0, 1.00=0, 5,000.00=0, 2,000.00=0, 1,800.00=0, 2,065.00=0, 1,200.00=0, 1,600.00=0, 7,657.00=1, 11,711.00=0, 10.00=0, 350.00=0, 2,200.00=0, 2,500.00=0, 380.00=0, 590.00=0, 500.00=0, 680.00=0, 450.00=0, 650.00=0, 420.00=0, 3,360.00=1, 3,185.00=0, 1,488.00=1}

| Sucursal = PIJUIAPAN: 2,000.00 {1,000.00=0, 1,500.00=1, 1.00=0, 5,000.00=0, 2,000.00=2, 1,800.00=0, 2,065.00=0, 1,200.00=0, 1,600.00=0, 7,657.00=1, 11,711.00=1, 10.00=0, 350.00=0, 2,200.00=0, 2,500.00=0, 380.00=0, 590.00=0, 500.00=0, 680.00=0, 450.00=0, 650.00=0, 420.00=0, 3,360.00=0, 3,185.00=0, 1,488.00=0}

| | | Actividad = COMERCIO AL POR MENOR DE ALIMENTOS BEBIDAS Y TABACO: 1,800.00 {1,000.00=0, 1,500.00=0, 1.00=0, 5,000.00=0, 2,000.00=0, 1,800.00=2, 2,065.00=0, 1,200.00=0, 1,600.00=0, 7,657.00=0, 11,711.00=0, 10.00=0, 350.00=0, 2,200.00=1, 2,500.00=0, 380.00=0, 590.00=0, 500.00=0, 680.00=0, 450.00=0, 650.00=0, 420.00=0, 3,360.00=0, 3,185.00=0, 1,488.00=0}

| | | Actividad = OTROS SERVICIOS DE APOYO A LOS NEGOCIOS: 380.00 {1,000.00=0, 1,500.00=0, 1.00=0, 5,000.00=0, 2,000.00=0, 1,800.00=0, 2,065.00=0, 1,200.00=0, 1,600.00=0, 7,657.00=0, 11,711.00=0, 10.00=0, 350.00=0, 2,200.00=0, 2,500.00=0, 380.00=1, 590.00=0, 500.00=1, 680.00=0, 450.00=0, 650.00=0, 420.00=0, 3,360.00=0, 3,185.00=0, 1,488.00=0}

Figura 7

Representación de árbol de decisión para generar conocimiento.

Conclusiones

Las raíces de la minería de datos son la inteligencia artificial y la estadística. La primera, establece soluciones algorítmicas a costes razonables, y la segunda se ocupa de generalizar los resultados obtenidos, trasladando los datos a situaciones más comunes. Los avances tecnológicos hacen que la colecta de datos avance y se generen nuevos problemas, esta situación hace que se tengan que utilizar tablas de datos más complejas y de gran tamaño.

Las técnicas de minería de datos para resolver problemas empresariales que implican el tratamiento de gran cantidad de información, son una medida que muchas entidades financieras deberían tomar en cuenta para la óptima implementación de las reglas dispuestas para la prevención de lavado de dinero, confiando en los pasos de este modelo para obtener resultados en sus procesos de desarrollo.

Sin embargo, muchas entidades financieras en México no han sido capaces de utilizar estructuralmente los procesos y técnicas que brinda la minería de datos, y varias de estas organizaciones utilizan metodologías y técnicas que hoy día ya no están dando resultados o bien muchas instituciones simplemente carecen de estos procesos. Las entidades financieras creen que estos posesos y metodologías les generarán grandes costos, pero en realidad, al adoptar estos modelos, reducirán sus costos enormemente, debido a que las regulaciones propuestas en el combate al lavado de dinero serán alcanzadas con el presupuesto y tiempo requerido.

El implementar y adoptar máquinas de aprendizaje, con una visión de integración y desarrollo empresarial, fortalecerá los procesos en las entidades financieras mexicanas, garantizando la detección de transacciones ilícitas.

Referencias

Bauer, Richard J. JR. (1994). *Genetic Algorithms and Investment Strategies*, John Wiley & Sons, Inc.

Davis, Lawrence. (1994). *Genetic Algorithms and Financial Applications*, in: Deboeck (ed., 1994), 133-141.

Kingdon, J. & Feldman, K. (1995). Genetic algorithms and applications to finance. *Applied Mathematical Finance*, Vol. 2, N° 2, Junio, Págs. 89-116.

Pereira, Robert. (1996). Selecting parameters for technical trading rules using genetic algorithms. *Journal of Applied Finance and Investment*, Volume 1, No. 3, July/August, 27-34.

Martin Marin, J.L. (1984). "El diagnóstico de las dificultades empresariales mediante modelos de mercado de capitales", *Boletín de Estudios Económicos*, N° 123, diciembre, Págs. 621-635.

Rich, E. (1994). *Inteligencia Artificial*, McGraw-Hill/Interamericana de España, segunda edición.

Sosa Sierra, M.C. (2007): "Inteligencia Artificial en la gestión financiera empresarial", *Boletín Pensamiento y Gestión*, No. 23. Colombia.

Ramírez Quintana, María José y Hernández Orallo, José (2003). *Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software*. España.

Serrano, C. y Martín del Brío, B. (1993): "Predicción de la crisis bancaria mediante el empleo de redes neuronales artificiales", *Revista Española de Financiación y Contabilidad*, Vol. XXII, N° 74, Págs. 153-176.

Virseda Benito, Fernando y Román Carrillo, Javier (sin año). *Minería de datos y aplicaciones*. Universidad Carlos III.