

Tipo de artículo: Artículo original
Temática: Soluciones Informáticas
Recibido: 11/06/2020 | Aceptado: 12/10/2020 | Publicado: 01/11/2020

La minería de datos y algunas de sus aplicaciones contextuales

Data mining and some of its contextual applications

Alberto Rodríguez Rodríguez^{1*}, Holger B. Delgado Lucas², Wilter Leonel Solórzano Álava³ Xiomara Lisbeth Anzules Ávila⁴

¹ Docente de la Universidad Estatal del Sur de Manabí, Jipijapa – Manabí – Ecuador. alberto.rodriguez@unesum.edu.ec

² Decano de la Facultad de Ciencias Técnicas de la Universidad Estatal del Sur de Manabí, Jipijapa – Manabí – Ecuador. holger.delgado@unesum.edu.ec

³ Universidad Estatal del Sur de Manabí, Jipijapa – Manabí – Ecuador. E-mail: solorzano-wilter7647@unesum.edu.ec

⁴ Universidad Estatal del Sur de Manabí, Jipijapa – Manabí – Ecuador. E-mail: anzules-xiomara8108@unesum.edu.ec

* Autor para correspondencia: alberto.rodriguez@unesum.edu.ec

Resumen

Este trabajo investiga una de las técnicas novedosas en la Programación en Enteros (PE), el método de Ramificación y Acotación combinado con los cortes de Gomory y Chvatal-Gomory. A través de los programas Programación Lineal en Enteros v.1.0 y del SPSS v.20.0 se lleva a cabo la implementación y comparación de distintas alternativas empleadas en el método de Ramificación y Corte para problemas pequeños, utilizando los cortes mencionados anteriormente y aplicados a diferentes niveles del árbol. Se hace una comparación con el software Mathematica v.9.0 para validar los resultados obtenidos. Fueron analizadas 12 estrategias para la resolución de problemas de PLE.

Palabras clave: aplicación; proceso; técnicas

Abstract

This research was based on the analysis of the different uses and applications of data mining and how the execution of each of the data mining techniques affects companies in today's world, the importance of this research lies with the advancement of technology and how companies have the need to evolve the analysis and processing of the data of each of their clients. The objective is to analyze each of the data mining processes in order to know what is the best technique that companies can implement and achieve an impact with incidence in today's world. Scientific research methods were used as the historical-logical method, it

was used to search the antecedents of data mining, the analysis-synthesis method, which was used to inquire between the different techniques that can be carried out in companies. It was possible to know each of the data mining processes and how these are important to execute in order to validate each of the data of its clients, workers and investors of a company that applies these data mining techniques.

Keywords: *application; process; techniques*

Introducción

En el mundo, la minería de datos es un proceso que se está desarrollando en la mayoría de empresas, busca encontrar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados. Usando una amplia variedad de técnicas, puede usar esta información para aumentar sus ingresos, reducir costos, mejorar las relaciones con sus clientes, reducir riesgos y más, con el objetivo de crear subconjuntos de datos (Content, 2019)

En el Ecuador la mayoría de empresas y corporaciones están empezando a utilizar los procesos de la minería de datos para poder buscar cada uno de sus patrones y utilizando cada una de estas técnicas, está aplicada en las industrias, agricultura medicina, la aplican investigadores para ver los datos meteorológicos y los estados del clima, y estos procesos contribuyen a mejorar en la toma de decisiones para la evolución e innovación de cada una de las empresas que los apliquen. (Camana, 2016)

Los procesos llevados a cabo por la minería de datos sirven como garantía para las empresas que lo aplican porque mejorarán sus ingresos, y pueden tener un presupuesto que les permita seguir teniendo esta tecnología a un alto nivel, también ayuda a las empresas a garantizar que sus datos no tienen riesgos de ser violados porque al monitorear constantemente la información del cliente, están protegidos con altos niveles de seguridad.

El objetivo de esta investigación se enfoca en analizar cada uno de los procesos de la minería de datos para de esta forma saber cuál es la mejor técnica que pueden implementar las empresas y así mejorar en las diferentes anomalías que puedan encontrarse. También se planteó conocer cada una de las aplicaciones de la minería de datos en el mundo actual como se utiliza y que beneficios aporta. El desarrollo del contexto de esta investigación permitirá el progreso conceptual y práctico de cada uno de los procesos de la minería de datos.

Materiales y métodos

La minería de datos es el proceso de búsqueda en grandes bases de datos para encontrar información útil que sirva para la toma de decisiones en las empresas, el software de minería de datos es uno de una serie de instrumentos

analíticos para el análisis de datos. Permite a los usuarios analizar los datos desde muchas dimensiones o ángulos, clasificar, y resumir las relaciones identificadas.(López , 2019)

Se puede entender la minería de datos o data mining como la tecnología y el software utilizados para encontrar patrones de comportamiento en la base de datos. La base básica para esto es que estos modelos ayudan en la toma de decisiones. Por ejemplo, esto podría ayudar a las empresas a comprender los patrones de comportamiento de sus clientes. Por lo tanto, esto facilitaría la implementación de estrategias para aumentar las ventas o reducir los costos.(López , 2019)

Etapas de la minería de datos

Dentro de un proceso de minería se encuentran cinco fases:

- **Objetivo y recolección de datos:** Lo primero de todo es enfocarse en qué tipo de información queremos obtener. Imaginemos el ejemplo de que un supermercado quiere saber a qué hora del día es donde hay más asistencia al cliente. Este sería el objetivo y la información que el comercio quiere obtener en este caso.
- **Procesamiento y gestión de los datos:** Una vez que conocemos los datos que queremos recopilar, nos ponemos a trabajar. Esta es quizás la fase más complicada del proceso. Requiere seleccionar la muestra representativa sobre la cual llevar a cabo el análisis. Una vez que se ha elegido la muestra, se debe analizar el tipo de variables o el modelo de regresión que se realizará en la muestra.
- **Selección del modelo:** Está estrechamente relacionado con la fase anterior. Se trata de crear un modelo o algoritmo que nos brinde el mejor resultado posible. Para esto, se debe realizar un análisis exhaustivo de las variables que se incluirán en el modelo. Esto se convierte en una tarea complicada ya que dependerá del tipo de información a analizar. Por lo tanto, los mineros de datos realizan diferentes pruebas del algoritmo, tales como: regresión lineal, árbol de decisión, series de tiempo, red neuronal, etc.
- **Análisis y revisión de resultados:** Básicamente es analizar los resultados para ver si dan una explicación lógica. Explicación que facilita la toma de decisiones basada en la información proporcionada por los resultados.
- **Actualización del modelo:** El último paso en el proceso sería actualizar el modelo. Es muy importante que se haga con el tiempo para que no quede obsoleto. Las variables del modelo podrían volverse insignificantes y, por lo tanto, se requiere un control periódico del mismo. (López , 2019)

Técnicas de la minería de datos

Las técnicas de minería de datos provienen de la inteligencia artificial y las estadísticas en sí. Estos son algoritmos que se aplican a un conjunto de datos para obtener resultados. Las técnicas más representativas utilizadas en la

minería de datos son las redes neuronales. Es un paradigma de aprendizaje y procesamiento automatizado que se inspira en la forma en que funciona el sistema nervioso de los animales. Es un sistema que permite que las neuronas se interconecten en una red (red neuronal) que proporciona colaboración para la producción de estímulos de salida.(Ventura, 2020)

Los ejemplos de redes neuronales incluyen:

- La perceptrón (tipo de red neuronal artificial)
- La perceptrón multicapas
- Los mapas autoorganizados

Redes Neuronales

Son un sistema de aprendizaje y procesamiento automático inspirado en la forma como funciona el sistema nervioso animal, que parte de la interconexión de neuronas que colaboran entre sí para generar un resultado.(Suárez, 2014)

Las redes neuronales son un modelo para encontrar esa combinación de parámetros y aplicarla al mismo tiempo. En su propio idioma, encontrar la mejor combinación es "entrenar" la red neuronal. Una red ya capacitada puede usarse para hacer predicciones o clasificaciones, es decir, para "aplicar" la combinación.(Guillermo , 2016)

Árbol 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 cada árbol hay un nodo inicial llamado raíz, que contiene toda la información.(Suárez, 2014)

Este método analítico que a través de una representación esquemática de las alternativas disponibles facilita una mejor toma de decisiones, especialmente cuando existen riesgos, costos, beneficios y múltiples opciones. El nombre se deriva de la apariencia del modelo en forma de árbol y su uso es amplio en el campo de la toma de decisiones bajo incertidumbre junto con otras herramientas como el análisis del punto de equilibrio.(López, 2019)

Se pueden seguir las siguientes técnicas para la toma de decisiones correctamente:

- Plantear problemas desde diferentes perspectivas.
- Analizar todas las posibles soluciones.
- Estimar los costes empresariales de cada decisión.
- Realizar unas decisiones razonadas con información real y existente de tu empresa.
- Analizar aquellas alternativas y probabilidades que den como resultado el éxito de tu negocio.

Agrupamiento (clustering)

Estos procesos de agrupamiento o también llamados clustering, buscan generar nuevos conjuntos a partir de los datos analizados, de modo que los individuos en el mismo grupo sean lo más similares posible y, simultáneamente, cada grupo sea lo más diferente posible. Estos métodos funcionan con la distancia entre cada elemento y el centro de cada uno de los grupos, asignando el individuo al grupo más cercano. (Suárez, 2014)

Según (Martinez, 2020) Se puede usar clustering para comprimir imágenes con pérdida de información. La compresión, en este ejemplo, se hace en el número de colores diferentes que se usan. Se supone que la imagen original utiliza una paleta de 255 colores. Para comprimir la imagen, se decide usar menos bits por pixel, es decir, usar menos colores. En la siguiente figura puedes ver cómo quedaría la foto original a medida que se usa menos y menos colores. Primero 10 colores, luego 3 y por último sólo 2. La técnica de clustering ayuda a decidir qué nuevos colores pueden representar mejor la imagen original cuando limitamos el número posible de colores a usar (K).

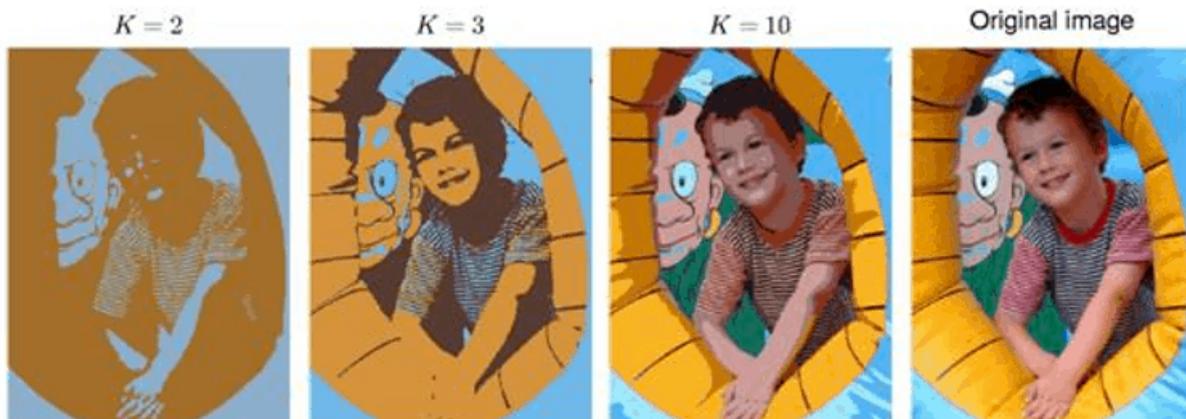


Figura 1.-Compresión de datos

Fuente: <https://rpubs.com/yujingma45/155921>

Aplicaciones de la minería de datos

En el mundo actual la minería de datos la podemos aplicar en diferentes campos de trabajo, a continuación, se presentarán sus principales aplicaciones:

- 1. Análisis de datos financieros:** se emplea tanto en el sector bancario como en el de las finanzas. Se busca proveer datos que aseguren que es posible practicar análisis sistemáticos en condiciones avanzadas y con garantías de fiabilidad. (Logicalis, 2014)

Algunos ejemplos son:

- Diseño y construcción de almacenes de datos para el análisis multidimensional de datos.
- Predicción de pago de préstamos y análisis de políticas de crédito de cliente.

- Clasificación y el agrupamiento de los clientes para la creación de ofertas personalizadas.
 - Detección de blanqueamiento de dinero y otros delitos financieros.
- 2. Industria minorista:** el sector retail recoge grandes cantidades de datos provenientes de las ventas, el historial de compra de los clientes o el transporte de mercancías. La cantidad de datos recogidos continúa expandiéndose rápidamente debido al aumento de la facilidad, disponibilidad y popularidad de la web y las transacciones online. La minería de datos con sus aplicaciones para la industria minorista ayuda a identificar patrones de compra de los clientes y tendencias. De esta forma, las empresas están en condiciones de proporcionar una mejor calidad de servicio al cliente, aumentando su satisfacción y facilitando su retención. (Logicalis, 2014)

Entre estas aplicaciones destacan las que permiten:

- El análisis multidimensional de las ventas, los clientes, los productos, el tiempo y la región. Los análisis de la eficacia de las campañas de ventas.
 - La recomendación personalizada de productos.
 - Las referencias cruzadas de artículos.
- 3. Industria de las telecomunicaciones:** en este sector, los datos son especialmente importantes para alcanzar una buena comprensión del negocio. La minería de datos y aplicaciones específicamente diseñadas para esta área, ayudan en la identificación de los patrones de telecomunicaciones, facilitan la detección de actividades fraudulentas y posibilitan el hacer un mejor uso de los recursos, mejorando la calidad del servicio. (Logicalis, 2014)

Entre las más ventajosas están:

- Análisis multidimensional de datos de telecomunicaciones.
 - Análisis de patrones fraudulentos.
 - Identificación de patrones inusuales, hábitos y tendencias.
 - Asociación multidimensional y análisis de patrones secuenciales.
- 4. Análisis de datos biológicos:** el campo de la biología es uno de los más beneficiados por los avances de la tecnología. La genómica, la proteómica, la genómica funcional y la minería de datos aplicada a la investigación de los seres vivos son sólo algunos ejemplos, una lista donde no hay que olvidarse de la bioinformática. (Logicalis, 2014)

La minería de datos con sus aplicaciones aporta una contribución importante para el análisis de datos biológicos:

- Integración semántica de las bases de datos genómicos y proteómicos heterogéneos distribuidos.

- Alineamiento, indexación, búsqueda de similitudes y análisis comparativo de múltiples secuencias de nucleótidos.
- Descubrimiento de patrones y análisis de redes genéticas.
- Identificación de patrones de proteínas estructurales.

Resultados y discusión

Se logró conocer cada uno de los procesos de la minería de datos y como estos son importantes ejecutarlos para así poder validar cada uno de los datos de sus clientes, trabajadores e inversionistas de una empresa que aplica estas técnicas de minería de datos, gracias a la aplicación de la data mining las industrias han podido reducir los riesgos de pérdida de información y además han logrado incrementar sus ingresos anuales en sus empresas.

Se analizó cada uno de los procesos fundamentales de la minería de datos, como podemos observar en la figura 2 , muestras estos procesos que debemos donde se denota que primero debemos establecer los objetivos y también delimitar hasta donde vamos a llegar con el cliente en el procesado de su información, en el procesado que es el punto principal y más importante de realizar porque aquí se van procesar y monitorear todos los datos del cliente para poderlos transformar en una base esta etapa lleva ente el 60 y 70 % de todo el proceso.

En el proceso de modelación o determinación se analizó los datos estadísticos para después representarlos en una gráfica, y de acuerdo a los objetivos planteados se tomará la mejor decisión sobre en qué herramienta o técnica se utilizará para procesar sus datos. En la etapa final o verificación de resultados se validó si los datos obtenidos son coherentes y acorde con los datos representados por el análisis estadístico y de visualización gráfica, aquí el cliente determinará si los datos obtenidos son novedosos y si le aportan un nuevo conocimiento que le permita razonar sus decisiones.

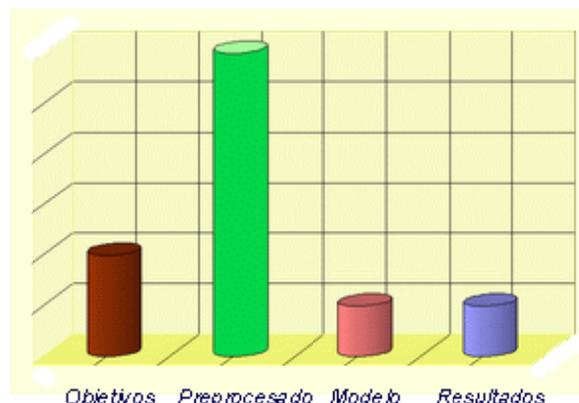


Figura 2. – *Procesos de minería de datos*

Fuente: https://www.sinnexus.com/business_intelligence/datamining.aspx

Conclusiones

- Para concluir cada una de estas etapas, procesos y técnicas de la minería de datos son muy importante sus aplicaciones, porque están aportan varios beneficios para el desarrollo de la industria comercial y también tecnológica para que cada empresa pueda obtener resultados apropiados y validos que contribuirán a la toma de decisiones.
- La minería de datos contribuye mucho para la evolución e innovación de las industrias de tal forma que permite el analizar cada uno de sus procesos y además estos generan nuevos conocimientos que contribuyen al desarrollo sostenible y productivo, ya sea en la educación, medicina, agricultura, telecomunicaciones, entre otros

Referencias

- Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica ESPOL*, 170-183.
- Content, R. (2019). *Rock Content*. Obtenido de rockcontent.com: <https://rockcontent.com/es/blog/mineria-de-datos/>
- Guillermo, J. (2016). *Xataka*. Obtenido de xataka.com: <https://www.xataka.com/robotica-e-ia/las-redes-neuronales-que-son-y-por-que-estan-volviendo>
- Logicalis. (2014). *Logicalis Architects of Change*. Obtenido de logicalis.com: <https://blog.es.logicalis.com/analytics/mineria-de-datos-aplicaciones-que-ya-son-una-realidad>
- López, J. (2019). *Economipedia*. Obtenido de economipedia.com: <https://economipedia.com/definiciones/mineria-de-datos.html>
- López, D. (2019). *Emprende Pyme*. Obtenido de emprendepyme.net: <https://www.emprendepyme.net/que-es-y-como-se-hace-un-arbol-de-decisiones.html>
- Martinez, J. (2020). *IArtificial*. Obtenido de iartificial.net: <https://iartificial.net/clustering-agrupamiento-kmeans-ejemplos-en-python/>
- Suárez, L. (2014). TÉCNICAS DE MINERÍA DE DATOS PARA LA DETECCIÓN Y PREVENCIÓN DEL LAVADO DE ACTIVOS Y LA FINANCIACIÓN DEL TERRORISMO (LA/FT). *Unidad de Información y Análisis Financiero (UIAF) - Colombia*.

Ventura, S. (2020). *Gestion.Org*. Obtenido de *gestion.org*: <https://www.gestion.org/tecnicas-de-mineria-de-datos/>