

Minería de datos en el análisis de tendencias políticas en redes sociales



Data mining in the analysis of political trends in social networks

Edwin León Plusas.¹, Esmeralda Proaño Arias.², Viena Muirragui Irrazábal.³ & Jaime Cajamarca Yunga.⁴

Recibido: 23-06-2019 / Revisado: 29-07-2019 / Aceptado: 09-08-2019 / Publicado: 10-09-2019

Abstract

DOI: <https://doi.org/10.33262/cienciadigital.v3i3.4.837>

Currently Social Networks are immersed in all fields of daily life, in areas such as engineering, architecture, health, economics, politics, etc., which is why they have become a major factor in the decision of the user, for this reason the present work is developed with the purpose of using Data Mining as a tool to transform the information collected from social networks on political trends. The investigation starts from the theoretical study of data mining with the definition of necessary concepts for the process, then the tool for the collection of information was designed, which was applied to users of social networks such as Facebook, Twitter, WhatsApp and Instagram, because it was studied about the political tendencies in these networks. Of the total population, through a non-probabilistic sampling for convenience, a total of 1016 people were selected for the application of the survey, with which it was possible to obtain the necessary information to perform the required analysis with the data mining technique and later in the data analysis it can be specified that 24.3% of the total respondents are of the center tendency, 14.6% are of the center right, 18.2% are of the center left, 9.5% are of the right, 15.1% are of the left and the 18.3% of the voters took as an option that they do not belong to any political tendency.

Keywords: Data Warehouse, data mining, politics, social networks.

¹ Universidad Estatal de Milagro, eleonp@unemi.edu.ec

² Universidad Tecnológica Israel, esmeraldap2007@gmail.com

³ Universidad Estatal de Milagro, vmuirragui1@unemi.edu.ec

⁴ AKROS Soluciones Tecnológicas, jaime.cajamarca@akroscorp.com

RESUMEN

Actualmente las Redes Sociales se encuentran inmersas en todos los campos del diario vivir, en áreas como ingeniería, arquitectura, salud, economía, política, etc., por lo que se han constituido en un factor de gran peso en la decisión del usuario, por tal motivo el presente trabajo se desarrolla con la finalidad de utilizar la Minería de Datos como herramienta para transformar la información recopilada de las redes sociales sobre las tendencias políticas. La investigación parte del estudio teórico de la minería de datos con la definición de conceptos necesarios para el proceso, posteriormente se diseñó la herramienta para la recopilación de información, la cual fue aplicada a los usuarios de redes sociales como Facebook, Twitter, WhatsApp e Instagram, debido a que se estudió acerca de las tendencias políticas en dichas redes. Del total de la población, mediante un muestreo no probabilístico por conveniencia, se seleccionó un total de 1016 personas para la aplicación de la encuesta, con la cual se pudo obtener la información necesaria para realizar el análisis requerido con la técnica de minería de datos y posteriormente en los análisis de datos se puede especificar que el 24.3% del total de encuestados son de tendencia de centro, 14.6% son de centro derecha, 18.2% son de centro izquierda, 9.5% son de derecha, 15.1% son de izquierda y el 18.3% de los votantes tomaron como opción que no pertenecen a ninguna tendencia política.

Palabras claves: Data Warehouse, minería de datos, política, redes sociales.

1. INTRODUCCIÓN

En los últimos años, la industria de las comunicaciones ha centrado su atención en sistemas que transportan datos a largas distancia. La industria de la informática y de las redes WAN (Wide Area Network) es un sector estable en constante desarrollo abarcando cada vez más y más actividades personales, profesionales y empresariales.

Tradicionalmente las personas disponían de su propia computadora, con la cual ejecutaban sus propias aplicaciones y administraban sus actividades e información de forma independiente conectada a una red de recursos, pronto surgió la necesidad de compartir esos recursos con el entorno social acoplándose a la tendencia de las comunicaciones globales para lograr los objetivos establecidos (Reyes, 2019).

El factor tecnológico disponible solo al entorno profesional, pronto llegó al sector personal cubriendo un entorno social muy diverso con la aplicación de las TIC (Tecnologías de Información y Comunicación) que engloban un amplio rango de las Redes Sociales (RRSS) disponibles, las cuales están siendo utilizados por personas sin diferenciar nivel social, género, ni edad.

Con estas aseveraciones se deduce que desde hace años atrás se ha evidenciado la necesidad de administrar la información que generan todo tipo de organizaciones, lo que exige llevar de manera ordenada datos que faciliten la toma de decisiones, por tal motivo se promueve la generación masiva de información que creen históricos de los movimientos que realizan las organizaciones, sin embargo actualmente en la mayoría de los casos este control se lleva de forma no muy desarrollada, por lo que los resultados no son los que se desean (Benalcázar, 2017).

Dentro de las organizaciones mundiales se encuentran las redes sociales las cuales indiscutiblemente juegan un papel importante en la vida social, económica y política de las personas, debido a que han permitido transformar la dinámica para la comunicación y adquisición de información, llegando a ser un factor de gran peso en la decisión del usuario despreciando la veracidad o falsedad de la información.

De acuerdo a ello, en la actualidad la técnica de minería de datos está siendo aplicada de forma muy notoria en diversas áreas, tales como el marketing, finanzas, medicina, salud pública, educación, entre otras, debido a que la aplicación de esta técnica sobre los datos provenientes de fuentes como las redes sociales pueden revelar patrones sobre individuos inmersos en el ambiente compartido y producir conocimiento que años atrás no era factible encontrar, pues la información era muy variada y compleja (Jaramillo, Cardona, & Fernández, 2015).

Por lo expuesto anteriormente, el presente trabajo de investigación tiene como finalidad analizar las redes sociales de mayor impacto en conjunto con los diferentes algoritmos de minería de datos que puedan utilizarse para generar conocimiento a partir de la información de dichas redes.

2. DESARROLLO

Aproximación al concepto de minería de datos

Inicialmente la minería de datos se puede definir como un proceso en el cual se analiza una gran cantidad de datos con la finalidad de descubrir nuevas relaciones, tendencias y patrones significativos (Pérez, 2007).

El análisis de datos toma el nombre de *Data Mining* (Minería de datos) gracias a la disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas cuyo objetivo es descubrir de manera automática el conocimiento contenido en una extensa base de datos.

Las técnicas que se utilizan en esta metodología tienen como meta descubrir patrones, perfiles y tendencias mediante el análisis de los datos haciendo uso de tecnologías de reconocimiento de patrones, lógica difusa, redes neuronales, algoritmos genéticos y otras técnicas especiales de análisis de datos.

EIS: Executive Information System

Es un sistema que permite a los ejecutivos tomar decisiones en tiempo real, el cual no solo consta del software y de hardware sino también de toda la infraestructura que permite tener la información completa al alcance de su mano representada de manera gráfica al estilo de hojas de cálculo (ventas semanales, inventarios, balances, etc.) (Nieto, 2016).

OLAP (On-Line Analysis Planning)

Se denominan sistemas OLAP (Procesamiento analítico en línea) a aquellos que deben analizar datos desde diferentes perspectivas y soportar requerimientos complejos de análisis en un volumen ingente de datos.

Este sistema se caracteriza por ser una técnica de estudio multidimensional de datos en el cual el usuario puede navegar por los mismos de modo asistido (Molina & García, 2006)

En este sistema la información es vista como cubos los cuales están formados por categorías descriptivas (dimensiones) y valores cuantitativos (medidas) (Ver figura 1 y 2).

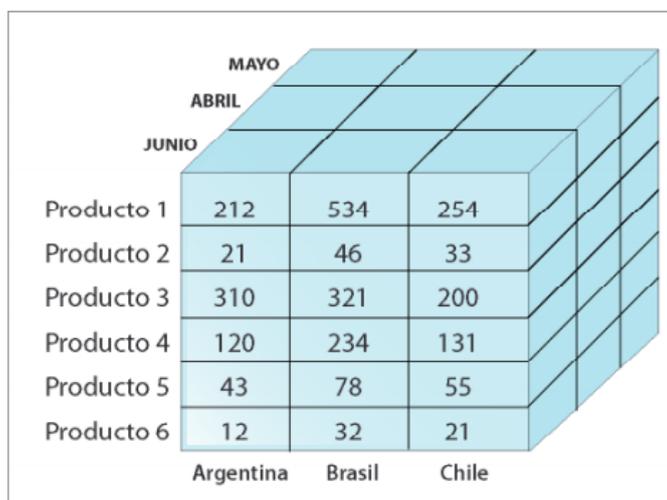


Figura 1. Vista de los datos en forma de cubo

Fuente: (Ibarra, 2006)

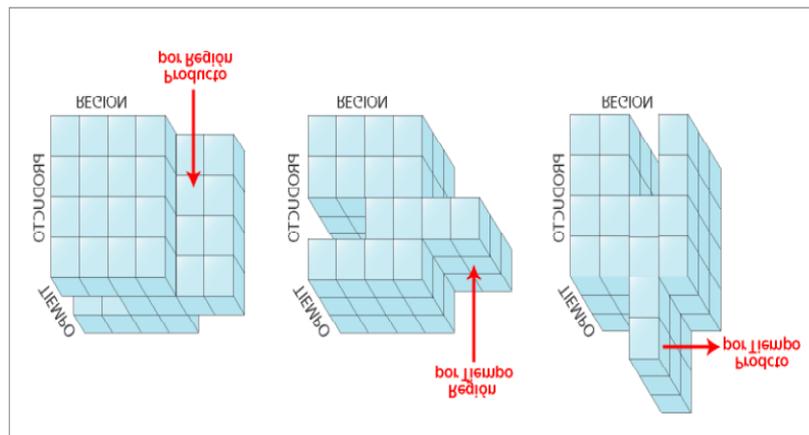


Figura 2 . Vista de los datos en forma de cubo
Fuente: (Ibarra, 2006)

Las ventajas de presentar de esta manera la información es que facilita a los usuarios diseñar consultas complejas, ordenar datos en un reporte, filtrar o disgregar datos en subconjuntos significativos y cambiar de datos resumidos a datos detallados.

Proceso de extracción del conocimiento a partir de datos (KDD)

El *Knowledge Discovery in Databases* es un proceso automático en el que se combinan análisis y descubrimiento. La tarea consiste en extraer patrones en forma de reglas o funciones para que el cliente los revise.

Las etapas básicas de este proceso son, reprocesar los datos, hacer minería de datos y presentar resultados.

Etapas del proceso KDD

- **Selección:** De acuerdo a los requerimientos del usuario y metas del proceso se crea un grupo de datos objetivo, escogiendo todo el conjunto de datos o una muestra de este, sobre el cual se efectuará el proceso de análisis.
- **Procesamiento / Clasificación:** En esta etapa se disgrega los datos de menor de calidad aplicando la técnica de remoción de datos ruidosos, datos nulos, datos duplicados. En esta fase en de suma importancia el trabajo a la par entre el usuario y el analista.
- **Conversión/reducción:** Se utilizan técnicas de reducción de dimensiones para disminuir el número de variables simplificando una tabla o base de datos vertical u horizontalmente. Las más utilizadas son agregaciones, compresión de datos, segmentación, histogramas, muestreo, etc.
- **Data mining (Minería de datos):** El objetivo de esta fase es descubrir patrones inimaginables e importantes aplicando las técnicas anteriormente expuestas.

- **Análisis de resultados /evaluación:** Interpretación de patrones descubiertos y generalmente se repite el ciclo en busca de nuevos hallazgos.

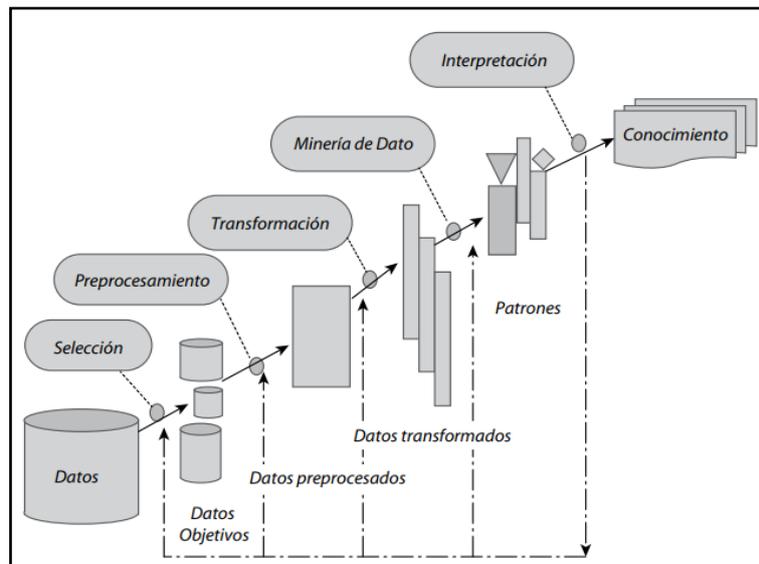


Figura 3. Etapas del proceso KDD

Fuente: (Timarán, Hernández, Caicedo, Hidalgo, & Alvarado, 2016)

Las redes sociales

El término *networking* se ha convertido en una palabra común, todas las personas van a una fiesta para gestionar su red de contactos, para un gran porcentaje de población la *World Wide Web* tiene como principal objetivo establecer contactos (Kadushin, 2012).

Las redes sociales han sido parte de nuestras vidas desde la época de los recolectores y cazadores, los individuos se generaban vínculos entre sí a través de sus relaciones unas con otras y su dependencia entre ellos. Las relaciones familiares y el parentesco son redes sociales, el barrio las ciudades son redes sociales. Fuera del parentesco familiar las personas de la sociedad moderna dependen unas de otras para el servicio de taxis, la venta del periódico el dentista, etc. Es decir, el parentesco en comunidades se forma a través de las necesidades y en torno a las redes sociales geográficamente dispersas es por ellos que se dice que llevamos milenios siendo *networkers*.

El *networking* es el uso de una de red de contactos para alcanzar diferentes objetivos personales, por lo tanto una red es simplemente un grupo de relaciones entre objetos que pueden ser organizaciones, personas, naciones en las cuales se comparte ideas de amistad, dinero, poder e incluso la enfermedad (Flores, Morán, & Rodríguez, 2015).

Hace solo algunos años (2003) se ha venido manejando el concepto de las redes sociales, pero como se describe en los párrafos anteriores esto no es del todo nuevo, sin embargo, lo que es relativamente nuevo es la manera de referirse a ellas describiéndolas

analizándolas y relacionándolas con instancias sociales más formales como organizaciones y gobiernos.

La red social viéndolo desde el lado informático se denomina como un servicio que brinda la oportunidad a los individuos de construir un perfil semipúblico o público dentro de un sistema delimitado, socializar con otros usuarios con los que comparten una conexión, ver y recorrer su historial de conexiones y consultar las de los demás usuarios del sistema (Flores, Morán, & Rodríguez, 2015).

3. METODOLOGÍA

La investigación parte del estudio teórico de la minería de datos con la definición de conceptos necesarios para el proceso, posteriormente se diseñó la herramienta para la recopilación de información, la cual fue aplicada a los usuarios de redes sociales como Facebook, Twitter, WhatsApp e Instagram, debido a que se estudió acerca de las tendencias políticas en dichas redes. Del total de la población, mediante un muestreo no probabilístico por conveniencia, se seleccionó un total de 1016 personas para la aplicación de la encuesta, con la cual se pudo obtener la información necesaria para realizar el análisis requerido con la técnica de minería de datos. Posteriormente para el procesamiento de información se utilizó la metodología Data Warehouse que es una base de datos corporativos encargada de integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista, con lo cual se pudo determinar que los usuarios si utilizan las redes sociales para compartir contenido político, siendo evidente que existe mayor afinidad por los partidos como Movimiento Centro Democrático, Partido Sociedad Patriótica, Movimiento Alianza País, entre otras.

4. RESULTADOS

En la presente investigación se tomó como población a todos los usuarios de redes sociales como Facebook, Twitter, WhatsApp e Instagram, debido a que se estudió acerca de las tendencias políticas en dichas redes. Del total de la población, mediante un muestreo no probabilístico por conveniencia, que de acuerdo a Otzen & Manterola (2017) permite seleccionar aquellos casos accesibles que acepten ser incluidos en el proceso o se encuentren disponibles en el tiempo o periodo de investigación, se seleccionó un total de 1016 personas para la aplicación de la encuesta, con la cual se pudo obtener la información necesaria para realizar el análisis requerido con la técnica de minería de datos.

Para el procesamiento de la información se utilizó la metodología DataWarehouse (DW), traducido como almacén de datos, misma que se trata de una base de datos corporativa que se caracteriza por integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista. La creación de un DataWarehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

Para el desarrollo del análisis y construcción del Data Warehouse se utilizó la Metodología Hefesto.

Es importante describir que también se utilizaron técnicas de limpieza y calidad de datos, procesos ETL y se realizó la carga inicial de información al Data Warehouse.

En este paso se debe realizar la Carga Inicial del DW, poblando el modelo construido, para lo cual se debe llevar adelante una serie de tareas básicas, tales como asegurar la limpieza y calidad de los datos, procesos ETL, etc.

En muchos casos, las tareas antes mencionadas tienen una lógica compleja, afortunadamente, en la actualidad existen muchas herramientas de software que se pueden emplear y que facilitan en gran parte el trabajo.

Una vez realizado la carga de información, se generó la Base de Datos a partir del Diagrama Físico del modelado, (Ver figura 4).

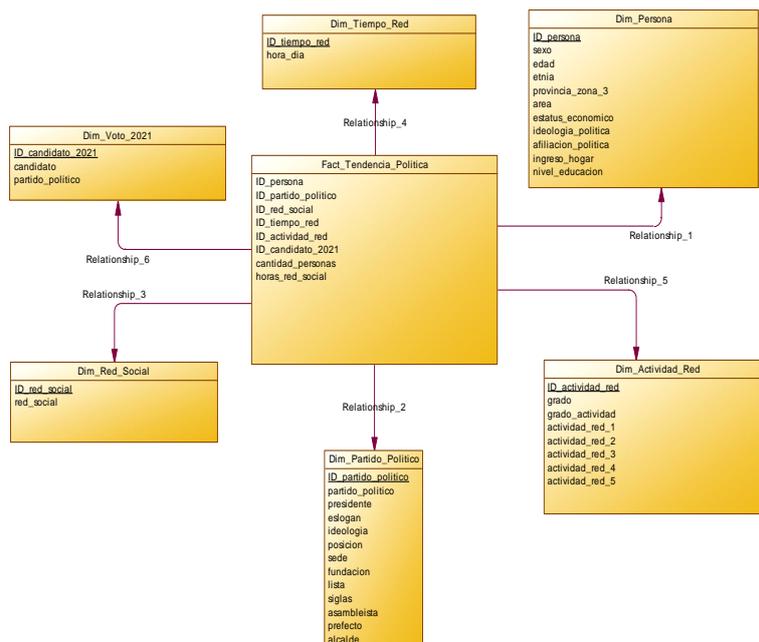


Figura 4. Diagrama físico del modelado.

Después de haber generado el modelo físico y haber realizado la carga de información, se muestra el grado de actividad de las personas en redes sociales, respecto a la ideología política que las personas indicaron tener. Además, se realiza la interpretación gráfica de

distintos modelos, cabe recalcar que las herramientas del cubo permiten girar las dimensiones permitiendo otra perspectiva de análisis.

Una vez realizado el proceso del análisis de información se puede representar las tendencias políticas de los usuarios encuestados, por lo cual, la figura 5, muestra información del grado de actividad que realizan las personas en las redes sociales, así como también la ideología política que las personas indicaron tener. Además, se realiza la interpretación gráfica de distintos modelos, cabe recalcar que las herramientas del cubo permiten girar las dimensiones permitiendo otra perspectiva de análisis.



Figura 5. Tendencias políticas

En las figuras 6 y 7, se muestra que la diferencia del promedio entre los votantes no es muy grande, por ello, en centro derecha e izquierda el porcentaje es de 1.3% de diferencia (Izquierda tiene el porcentaje más elevado), mientras que la tendencia de centro tiene un porcentaje 15.5% más que centro izquierda.

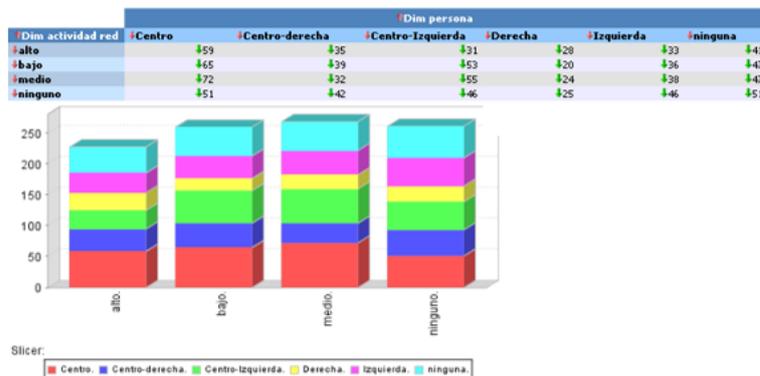


Figura 6. Interpretación gráfica apilada de Dim personas

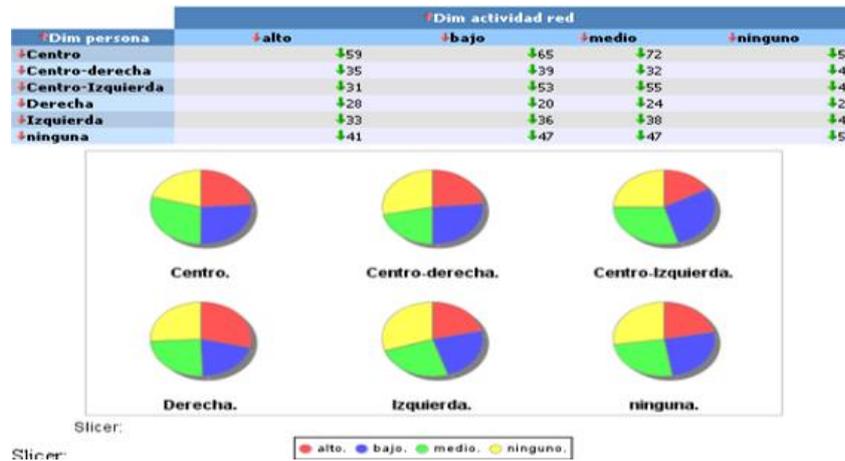


Figura 7. Tabulación de datos Dim actividad red

5. CONCLUSIONES

- En el estudio se puede evidenciar que las redes sociales han llegado a la sociedad para quedarse, pues permiten mantener la comunicación entre personas de diferentes culturas, etnias, creencias, costumbres, entre otros, extendiendo sus campos y llegando a todos los segmentos del mundo, ya que miles de usuarios expresan sus sentimientos, conocimientos y afinidades a través de esta información que puede ser aprovechada para proyecciones en diversos campos como la política.
- La forma de agrupación de la información en la minería de datos permite observar diferentes perspectivas a las bases de datos, por ejemplo, una agrupación de la base de datos de tendencias políticas podría exponer todos los partidos políticos que existen dentro de las regiones, mientras que otro grupo podría mostrar los partidos políticos a los que los usuarios tienen mayor afinidad.
- Con la metodología Data Warehouse que es una base de datos corporativos encargada de integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista, se pudo determinar que los usuarios si utilizan las redes sociales para compartir contenido político, siendo evidente que existe mayor afinidad por los partidos como Movimiento Centro Democrático, Partido Sociedad Patriótica, Movimiento Alianza País, entre otras.
- De acuerdo a los análisis de datos el 24.3% del total de encuestados son de tendencia de centro, 14.6% son de centro derecha, 18.2% son de centro izquierda, 9.5% son de derecha, 15.1% son de izquierda y el 18.3% de los votantes tomaron como opción que no pertenecen a ninguna tendencia política.

6. BIBLIOGRAFÍA

- Benalcázar, J. (2017). *Análisis comparativo de metodologías de minería de datos y su aplicabilidad a la industria de servicios*. Quito: Universidad de las Américas. Recuperado el mayo de 2018, de <http://dspace.udla.edu.ec/bitstream/33000/7547/6/UDLA-EC-TMGSTI-2017-11.pdf>
- Flores, J., Morán, J., & Rodríguez, J. (2015). *Las redes sociales*. Lima: Universidad de San Martín de Porres.
- Ibarra, M. (2006). *Procesamiento Analítico en Línea (OLAP)*. Universidad Nacional del Nordeste. Recuperado el mayo de 2018, de <http://exa.unne.edu.ar/informatica/SO/OLAPMonog.pdf>
- Jaramillo, S., Cardona, S., & Fernández, A. (2015). Minería de datos sobre streams de redes sociales, una herramienta al servicio de la Bibliotecología. *Información, cultura y sociedad*, 33, 63-74. Recuperado el mayo de 2018, de <http://www.redalyc.org/pdf/2630/263042678005.pdf>
- Kadushin, C. (2012). *Comprender las redes sociales*. Madrid: Centro de investigaciones sociológicas.
- Martínez, C. (2012). *Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión en entel*. Santiago de Chile: Universidad de Chile. Recuperado el mayo de 2018, de http://repositorio.uchile.cl/bitstream/handle/2250/112065/cf-martinez_ca.pdf?sequence=1
- Molina, J., & García, J. (2006). *Aplicaciones prácticas utilizando Microsoft Excel y weka*. Madrid, España: Universal Carlos Tercero de Madrid. Recuperado el mayo de 2018, de http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf
- Nieto, M. (2016). *Plan de negocios para la creación de una empresa de servicios de minería de datos en Colombia*. Bogotá: UNIR. Recuperado el mayo de 2018, de

<https://reunir.unir.net/bitstream/handle/123456789/4838/NIETO%20PATARROYO%2C%20MARCO%20ANTONIO.pdf>

- Pérez, C. (2007). *Minería de datos técnicas y herramientas*. Madrid, España: Paraninfo. S.A. Recuperado el mayo de 2018, de https://books.google.com.ec/books?id=wz-D_8uPFCEC&printsec=frontcover&hl=es#v=onepage&q&f=false
- Reyes, N. S. (2019). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. *Revista Ibérica de Sistemas e Tecnologías de Información*, 258-266.
- Timarán, S. R., Hernández, I., Caicedo, S. J., Hidalgo, A., & Alvarado, J. (2016). *El proceso de descubrimiento de conocimiento en base de datos*. Bogotá: Universidad Cooperativa de Colombia. Recuperado el mayo de 2018, de ediciones.ucc.edu.co/index.php/ucc/catalog/download/36/40/230-1

PARA CITAR EL ARTÍCULO INDEXADO.

León Pluas, E., Proaño Arias, E., Muirragui Irrazábal, V., & Cajamarca Yunga, J. (2019). Minería de datos en el análisis de tendencias políticas en redes sociales. *Ciencia Digital*, 3(3.4.), 91-103. <https://doi.org/10.33262/cienciadigital.v3i3.4.837>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Ciencia Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Ciencia Digital**.

