

S.E.P

INSTITUTO TECNOLÓGICO DE ORIZABA

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

MAESTRIA EN INGENIERÍA ADMINISTRATIVA

“MINERÍA DE DATOS, TEXTOS Y DE SENTIMIENTOS”

PRESENTA:

L.I. LOURDES GARCÍA MONTERO

NOMBRE DEL CATEDRÁTICO:

DR. FERNANDO AGUIRRE Y HERNÁNDEZ

Orizaba, Ver

Junio 2017

CONTENIDO

Introducción	3
Historia de las bases de datos	3
Big data dentro de las organizaciones	4
Minería de datos	5
Objetivo de la minería de datos	6
Funcionamiento de las minerías de datos	7
Minerías de textos	8
Aplicaciones de las minerías de textos	9
Minería de sentimientos	9
Ética en el uso de las minerías	10
Referencias bibliográficas	11
Agradecimientos	12

INTRODUCCIÓN

Sin lugar a dudas el valor máspreciado con el que la raza humana cuenta es el conocimiento, el cual a través de los años se ha ido almacenando de diferentes maneras.

Hoy en día este conocimiento se puede conocer con el nombre de datos o bien información, la cual ya no es solamente almacenada, si no analizada para la obtención de resultados.

Es así que se nacen los conceptos de minería de datos, de textos y de sentimientos, en un afán por pronosticar el futuro y facilitar la toma de decisiones.

Los concepto de minería de datos, de textos y de sentimientos son tópicos que aparecen en la sociedad actual debido a la existencia de grandes cantidades de información, este concepto se ha ido desarrollando en diferentes rubros de la vida cotidiana, tales como la educación, la salud, biología, análisis de mercados, telecomunicaciones, solo por mencionar algunas.

Son procesos informáticos que identifican patrones de comportamiento para poder realizar un pronóstico de las acciones futuras.

HISTORIA DE LAS BASES DE DATOS

Gracias a la especialización de los individuos a través de los años, surgió la necesidad de resguardar la información que estos producían, tales como investigaciones o títulos de propiedad, información básica de la población incluidas las actas de defunción, nacimiento y matrimonio.

Con el paso del tiempo, la humanidad no se conformó con el resguardo físico de dicha información, por lo que fueron desarrollando de manera paulatina dispositivos electrónicos que facilitarán el manejo de los datos, así como una gran cantidad de almacenamiento.

Durante el siglo XX se registraron diversas tecnologías como las fotoeléctricas, la aparición de las cintas magnéticas, las primeras computadoras como la UNIVAC, pero fue hasta finales de la década de los años 60's cuando se publica un artículo que define el modelo de bases de datos como tal.

Hoy en día se manejan grandes bases de datos capaces de generar, recopilar y procesar grandes cantidades de información, estas características se maximizan de manera paulatina convirtiéndose cada vez más fuertes (Martínez, E., 2000)

BIG DATA DENTRO DE LAS ORGANIZACIONES

Las organizaciones utilizan con mayor frecuencia este término gracias a diversos factores o medios de almacenamiento tales como la denominada “nube”, el Big data utiliza herramientas como: la minería de textos, minería de datos y la minería de sentimientos con el fin de facilitar la toma de decisiones, haciendo una organización cada vez más eficiente.

Gracias a la avanzada tecnología, se ha ido incrementado el tamaño del almacenamiento, así como los costos que le generan a la organización el uso de estos elementos.

Las organizaciones tienen como objetivos principales ser más eficientes y obtener mayores rendimientos, por lo que adicionan al análisis de sus datos, técnicas estructuradas o convencionales y técnicas no estructuradas o semiestructuradas,

como las minerías antes mencionadas, que adquieren su información de fotografías, audio, video, libros, entre otras.

Que sean capaces de brindar conclusiones de datos analizados con la mayor rapidez posible, todo este con el fin de obtener una ventaja competitiva, los datos que generalmente se manejan en este sentido son los siguientes:

- **Datos de una empresa tradicional** (transaccional). Información genérica de clientes, proveedores, suministros, etc.
- **Datos generados por sensores de datos.** Generalmente son sensores integrados a diferentes equipos desde su elaboración, también se denominan medidores inteligentes, inclusive chips
- **Datos sociales.** Se encuentran englobadas todas las redes sociales que brinda internet, siendo las más conocidas twitter o Facebook, y elementos tales como la fotografía y el audio.
- **Datos generados por movilidad.** Generalmente proporcionados por un equipo de tecnología móvil , pudiendo ser Tablet, Smartphone, o por medio de Skype.(Joyanes, L., 2013)

MINERÍA DE DATOS

Nace gracias al manejo constante de tecnologías de la información, |existen diferentes autores que se encargan de describirla, por lo que se mencionarán las definiciones más que se han identificado como más significativas.

“La minería de datos puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos”(Pérez, C. & Santín, D., 2008)

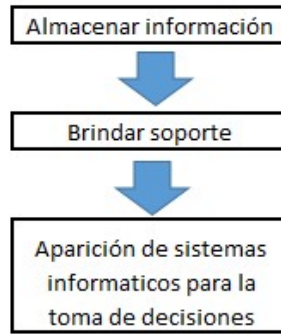
La minería de datos se considera como una combinación de técnicas semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos y que pueda resultar más provechosa que los propios datos de partida. (Martínez, E., 2000)

OBJETIVOS DE LA MINERÍA DE DATOS

Debido a la existencia de enormes cantidades de información y al gran avance de las denominadas tecnologías de la información se han generado técnicas especializadas para su análisis y control.

En estas técnicas se proponen realizar una identificación precisa de la información contenida, es decir no solo de manera textual, de lo contrario descubren patrones y tendencias de información, por medio de diferentes herramientas como: algoritmos, redes neuronales , análisis de datos, por mencionar algunas.(Pérez, C. & Santín, D., 2008)

Los datos y el manejo de la información han atravesado por diferentes fases a lo largo de la historia llegando finalmente a la minería de datos, por lo que es necesario considerar , la siguiente representación para brindar una idea más clara al respecto:

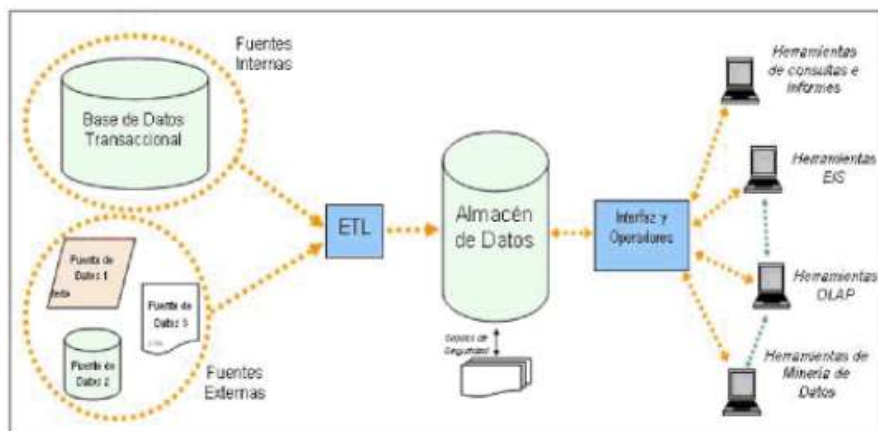


Fuente: Elaboración propia

FUNCIONAMIENTO DE LAS MINERÍAS DE DATOS

Esta técnica identifica y extrae patrones que “comprenden” de manera adecuada a los datos e información logrando establecer un pronóstico del comportamiento

Un ejemplo claro de la minería de datos es el denominado ETL (Extraction, Transformation, Load) dentro de sus fases se encuentra la lectura de datos, incorporación , transformación, mantenimiento y planificación de datos, (Pérez, C. & Santín, D., 2008) este modelo se presenta de la siguiente manera:



Fuente: (Pérez, C. & Santín, D., 2008)

MINERIAS DE TEXTOS

Esta técnica se basa en adquirir o encontrar información útil de un determinado rubro proveniente de una determinada cantidad de información generalmente se encuentran adicionadas en softwares especializados

Esta técnica se avoca en el descubrimiento de patrones, tendencias y asociaciones entre la diversidad de información existente. La minería de texto abarcan los siguientes rubros:

- ❖ Categorización de textos
- ❖ Clasificación de textos
- ❖ Generación de agrupamientos
- ❖ Descubrimiento de asociaciones
- ❖ Detección de desviaciones
- ❖ Análisis de tendencias
- ❖ Construcción de resúmenes (Montes, M., s/f)

El proceso de minería de texto se encuentra dividido en dos etapas:

- ❖ La primera denominada de procesamiento. Donde se realiza una representación de los textos en forma de estructuras.
- ❖ Y otra llamada de descubrimiento. Donde se describen los objetivos que se intentan descubrir tales como patrones o tendencias (Tan, 1999)

Generalmente los sistemas que son considerados como minería de textos cuentan con representaciones sencillas e identificación de algunas palabras consideradas clave lo que facilita el análisis e interpretación, por lo que se puede decir que sus representaciones se encuentran delimitadas por las tendencias establecidas.

APLICACIONES DE LA MINERÍA DE TEXTOS

La minería de textos se encuentra aplicable a diversos rubros, los más destacados son:

- ❖ **Aspectos de seguridad.** Se utiliza para dar seguimiento a las diferentes fuentes de información, así como para un estudio más detallado de posible texto cifrado.
- ❖ **Aspectos biomédicos.** Es un aspecto de investigación recién desarrollado donde se analizan el procesamiento del lenguaje natural, la informática, la bioinformática, inclusive la lingüística computacional.
- ❖ **Marketing.** Se expone el hecho de analizar la relación que las organizaciones poseen con los clientes, para evitar la pérdida de los mismos.
- ❖ **Aplicaciones académicas.** En muchos casos, al realizar la publicación de diversos artículos se solicita la indexación, esto se refiere a que exista una determinada cantidad y calidad de lados, establecida de forma muy específica.

MINERÍA DE SENTIMIENTOS

La ya conocida como minería de sentimientos es una derivación de la minería de textos, en este caso se intenta identificar las intenciones de los clientes o de los

usuarios en los diferentes medios, por ejemplo el correo electrónico, el Facebook, inclusive los diferentes buscadores con los que cuenta la red.

El proceso básico de la minería de datos se realiza de la siguiente manera:

- Se intenta entender o relacionar el texto con un rubro en específico, por ejemplo con un giro en especial, restaurants, o agencias de viajes.
- Posterior a ello se identifica la intención del cliente, colocando una etiqueta a las frases que lo integran como positivo o negativo, por ejemplo si en las bases de datos se encuentra clasificada la palabra “genial” como buena se adiciona una etiqueta positiva, caso contrario, si se encuentra la palabra “deficiente” clasificada como mala se adicionará una etiqueta negativa.

A través del análisis de estas etiquetas la organización tiene mayor conocimiento sobre las opiniones, y aflicciones de sus consumidores. (Pérez, S., 2016)

ÉTICA EN EL USO DE LAS MINERÍAS

La información que es manejada en algunas bases de datos es proveniente de usuarios, que en muchas ocasiones no se encuentran con pleno conocimiento de que diferentes organismos la usaran para los fines que crean convenientes.

Las compras electrónicas, los pagos, llamadas telefónicas entre otros son información que se genera día con día y que se va registrando de manera adecuada, la sociedad no se encuentra concientizada de la gran cantidad de información que genera y los peligros que pueden causar (Garriga, A., 2004)

Gracias a esta información la minería de datos muestra el tipo de influencia que se puede establecer con la sociedad para obtener determinados beneficios.

Sin embargo no todo es en contra de estas minerías, esta tecnología apoya a las organizaciones para hacer más eficientes a sus colaboradores, analizando sus comportamientos y habilidades.

Incluso mediante el seguimiento de los navegadores o correo electrónico se puede indagar si un colaborador se encuentra en la búsqueda de nuevas oportunidades de trabajo, o si presenta alguna conducta que lesione las actividades de la empresa misma inclusive el terrorismo.(Franganillo, J., 2010)

REFERENCIAS BIBLIOGRÁFICAS

Franganillo, J. (2010). Implicaciones éticas de la minería de datos. *Anuario thinkEPI*.

Garriga, A. (2004). Tratamiento de datos personales y derechos fundamentales.

Joyanes, L. (2013). *Big Data: Análisis de los grandes volúmenes de datos en organizaciones* (Primera). México D.F.: Alfaomega.

Martínez, E. (2000). *Minería de datos, una herramienta para la toma de decisiones*. UNAM. Recuperado a partir de <http://132.248.9.195/pd2001/287820/Index.html>

Montes, M. (s/f). *Minería de texto: Un nuevo reto computacional*. México D.F.: IPN. Recuperado a partir de <http://ccc.inaoep.mx/~mmontesg/publicaciones/2001/MineriaTexto-md01.pdf>

Pérez, C., & Santín, D. (2008). *Minería de datos, técnicas y herramientas* (Ediciones Paraninfo S.A.). España. Recuperado a partir de <https://books.google.com.mx/books?id=wz->

D_8uPFCEC&printsec=frontcover&dq=miner%C3%ADa+de+datos&hl=es-419&sa=X&sqi=2&redir_esc=y#v=onepage&q=miner%C3%ADa%20de%20datos&f=false

Pérez, S. (2016). Análisis de sentimientos. *Universidad Central de Venezuela*.

Tan. (1999). The state of the art and challenges, Proc. of the Workshop Knowledge Discovery from advanced Databases. *text mining*.

AGRADECIMIENTOS

Quiero| agradecer a Dios por permitirme un día más de vida, a mi esposo que emprende esta nueva aventura conmigo brindándome siempre su apoyo incondicional, al Instituto Tecnológico de Orizaba por darme una excelente formación académica en nivel licenciatura y ahora abrirme nuevamente las puertas en el área de postgrado, al Concejo Nacional de Ciencia y Tecnología (CONACYT) por ayudarme a cumplir mis sueños, al Dr. Fernando Aguirre y Hernández por abrirme los ojos ante un mundo lleno de posibilidades y caminos vagamente explorados.