

2017

MINERÍAS DE DATOS, TEXTOS Y SENTIMIENTOS

FUNDAMENTOS DE INGENIERÍA ADMINISTRATIVA
MARIELA DENISSE REBOLLO ALTAMIRA

INSTITUTO TECNOLÓGICO DE ORIZABA | MAESTRÍA EN INGENIERÍA ADMINISTRATIVA

USANDO LA TECNOLOGÍA A NUESTRO FAVOR

INTRODUCCIÓN

(Riquelme, Ruiz, & Gilbert, 2006) exponen que la revolución digital ha hecho posible que la información digitalizada sea fácil de capturar, procesar, almacenar, distribuir, y transmitir. Con el importante progreso en informática y en las tecnologías relacionadas y la expansión de su uso en diferentes aspectos de la vida, se continúa recogiendo y almacenando en bases de datos gran cantidad de información.

Por ello, es importante que las organizaciones utilicen este tipo de tecnologías a su favor, para poder enfrentar los desafíos que hay en la actualidad, como el mantener relaciones duraderas con los clientes. (Braga, Valencia, & Carvajal, 2009) declaran que el mayor desafío que las organizaciones pueden enfrentar hoy es el de mantener una cartera de clientes lucrativa, ya no se trata de organizar la producción, reducción de costos o la atención que se brinda, que aunque son condiciones necesarias ya no son suficientes para vencer en el mercado global y altamente competitivo.

Los autores sugieren que es mediante el conocimiento adquirido sobre los clientes, que las organizaciones deben ser capaces de interpretar sus objetivos, expectativas y deseos. Esto se puede conseguir, argumentan (Braga et al., 2009) por la “minería de datos” (data mining), o por “minería de datos dirigida al cliente” (customer centric data mining), que es una colección de técnicas y métodos que facilitan la adquisición y retención de la parte del mercado que cabe a una empresa (market share). Las metas de la buena atención y reducción de costos también valen para las organizaciones no lucrativas, gubernamentales o no. El negocio que conoce a sus clientes va a servirlos aún mejor.

El presente artículo explicará en que consiste la minería de datos, de textos y de sentimientos y como las organizaciones pueden utilizar este tipo de tecnologías, no únicamente para ganar ventaja competitiva frente a las demás organizaciones, sino también para conseguir un mejor enfoque en sus productos y/o servicios de la información obtenida de sus consumidores y clientes.

MINERÍAS DE DATOS

(Braga et al., 2009) explican que la minería de datos provee un método automático para descubrir patrones en datos, sin la tendenciosidad y la limitación de un análisis sustentado meramente en la intuición humana.

Así mismo explican que la minería de datos dirigida al cliente provee el conocimiento de las características y el comportamiento de los clientes. Dado que, retener a los clientes cuesta menos que adquirir nuevos.

Los autores explican que la minería de datos comprende un conjunto de técnicas para la descripción y predicción a partir de grandes masas de datos. Por este motivo, ella está generalmente asociada a bases de datos especialmente denominadas “data warehouse”. Estas bases de datos permiten la integración rápida de datos oriundos de diferentes fuentes.

(Joyanes, 2016) argumenta que la minería de datos se refiere al proceso de buscar información valiosa del negocio en una base de datos, data warehouse o data mart. La minería de datos puede realizar dos operaciones básicas:

- Predecir tendencias y comportamientos.
- Identificación de patrones desconocidos con anterioridad. Las aplicaciones normales de Bussines Intelligence normalmente proporcionan a los usuarios una visión de lo que ha sucedido, la minería de datos ayuda a explicar que está sucediendo y predice lo que sucederá en el futuro.

La minería de datos es un proceso que utiliza técnicas estadísticas, matemáticas, inteligencia artificial y de aprendizaje de máquinas para extraer e identificar información útil que convierte en conocimiento a partir de grandes bases de datos, data warehouse o data mart.

(Pérez, 2007) define a la minería de datos como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos.

(Riquelme et al., 2006) manifiesta que el descubrimiento de conocimiento en bases de datos (KDD) se define como el proceso de identificar patrones significativos en los datos que sean válidos, novedosos, potencialmente útiles y comprensibles para un usuario, y este proceso es interactivo e iterativo conteniendo los siguientes pasos:

1. Comprender el dominio de aplicación: este paso incluye el conocimiento relevante previo y las metas de la aplicación.
2. Extraer la base de datos objetivo: recogida de los datos, evaluar la calidad de los datos y utilizar análisis exploratorio de los datos para familiarizarse con ellos.
3. Preparar los datos: incluye limpieza, transformación, integración y reducción de datos. Se intenta mejorar la calidad de los datos a la vez que disminuir el tiempo requerido por el algoritmo de aprendizaje aplicado posteriormente.
4. Minería de datos: como se ha señalado anteriormente, este es la fase fundamental del proceso. Está constituido por una o más de las siguientes funciones, clasificación, regresión, clustering, resumen, recuperación de imágenes, extracción de reglas, etc.
5. Interpretación: explicar los patrones descubiertos, así como la posibilidad de visualizarlos.
6. Utilizar el conocimiento descubierto: hacer uso del modelo creado

Como hemos podido visualizar la fase fundamental del KDD es la minería de datos, por lo tanto, a continuación, explicaremos en qué consisten sus funciones:

- Clasificación: clasifica un dato dentro de las clases categóricas predefinidas.
- Regresión: el propósito de este modelo es hacer corresponder un dato con un valor real de una variable.
- Clustering: se refiere a la agrupación de registros, observaciones, o casos en clases de objetos similares. Un clúster es una colección de registros que son similares entre sí y distintos a los registros de otro clúster.

Hecho por: L.A Mariela Denisse Rebollo Altamira

Asesor: Dr. Fernando Aguirre y Hernández

- Generación de reglas: aquí se extraen o generan reglas de los datos. Estas reglas hacen referencia al descubrimiento de relaciones de asociación y dependencias funcionales entre los diferentes atributos.
- Resumen o sumarización: estos modelos proporcionan una descripción compacta de un subconjunto de datos.
- Análisis de secuencias: se modelan patrones secuenciales, como análisis de series temporales, secuencias de genes, etc. El objetivo es modelar los estados del proceso, o extraer e informar la desviación y tendencias en el tiempo.

APLICACIONES DE LA MINERÍA DE DATOS

(Riquelme et al., 2006) menciona que algunas de las tareas importantes de la minería de datos incluye la identificación de aplicaciones para las técnicas existentes, y desarrollar nuevas técnicas para dominios tradicionales o de nueva aplicación, como el comercio electrónico y la bioinformática.

Las áreas donde la minería de datos puede aplicarse, es prácticamente en todas las actividades humanas que generen datos como:

- Comercio y banca: segmentación de clientes, previsión de ventas, análisis de riesgo.
- Medicina y Farmacia: diagnóstico de enfermedades y la efectividad de los tratamientos.
- Seguridad y detección de fraude: reconocimiento facial, identificaciones biométricas, accesos a redes no permitidos, etc.
- Recuperación de información no numérica: minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.
- Astronomía: identificación de nuevas estrellas y galaxias.
- Geología, minería, agricultura y pesca: identificación de áreas de uso para distintos cultivos o de pesca o de explotación minera en bases de datos de imágenes de satélites.
- Ciencias Ambientales: identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales para mejorar su observación, gestión y/o control.
- Ciencias Sociales: Estudio de los flujos de la opinión pública. Planificación de ciudades: identificar barrios con conflicto en función de valores sociodemográficos.

MINERÍAS DE TEXTOS

(Matallana & Delgado, 2010) consideran que la minería de textos (text mining) una forma particular de la minería de datos que permite la extracción de conocimiento de grandes repositorios de información, estructurada o no, y en forma de texto. El objetivo es similar al de la minería de datos, descubrir patrones de comportamiento no visibles y nuevo conocimiento dentro de una colección documental.

La minería de textos aplica técnicas matemáticas y estadísticas, así como el análisis semántico del texto. La minería de textos es el proceso de aplicación de métodos automáticos para analizar y estructurar datos de texto con el objetivo de crear un conocimiento útil a partir de información estructurada y no estructurada.

La minería de textos, de acuerdo a estos autores, se enfoca al descubrimiento de patrones interesantes y nuevos conocimientos en un conjunto de textos, su objetivo es descubrir nuevas tendencias, desviaciones y asociaciones dentro de grandes volúmenes de información textual.

(Joyanes, 2016) explica que la minería de textos busca, mina y descubre texto en documentos de todo tipo, también es llamada minería de datos de textos, el autor argumenta que un sentido práctico es el proceso de deducir información de alta calidad a partir de un texto determinado.

El análisis de texto trata de encontrar patrones dentro de un conjunto de textos que facilite una mejor toma de decisiones, su objetivo es la mejora en la toma de decisiones. El análisis de textos pretende capturar datos no estructurados, procesados y crear a partir de ellos datos estructurados que puedan ser utilizados en los procesos de análisis y de creación de informes.

(Joyanes, 2016) argumenta que los datos de textos tienen un gran potencial de impacto en casi todas las organizaciones y empresas así como en la industria. El aprendizaje de métodos que permiten capturas, análisis gramaticales (parsing), y el análisis de texto final es crítico para las organizaciones.

MINERÍAS DE SENTIMIENTOS

(Joyanes, 2016) argumenta que la minería de sentimientos, se centra en el análisis de sentimientos y opiniones presentes en mensajes de texto y otros formatos de medio, y permiten descubrir a opinión o el sentimiento incrustado, por ejemplo, en mensajes de textos, en posts de twitter, etc. de un beneficio tangible para los accionistas y trabajadores que la integran.

La minería de opinión o de sentimiento como se le conoce, se refiere al procesamiento del lenguaje natural, lingüística computacional y analítica de texto para identificar y extraer información subjetiva de fuentes de materiales.

El análisis de sentimientos clásico ha sufrido un cambio espectacular desde la implantación del web 2.0 y el creciente uso de blogs y redes sociales. Una aplicación web que mide el análisis de sentimientos es "twitter sentiment". El análisis de sentimientos es en la actualidad un uso popular del análisis de textos, examina y obtiene la dirección general de la opinión a través de un número grande de personas que proporciona información sobre lo que el mercado está diciendo, pensando y sintiendo acerca de una organización o persona. El análisis de sentimiento utiliza los datos de los sitios de social media.

Desde la perspectiva de una organización o empresa, el análisis de sentimientos permite analizar de modo rápido y eficiente qué se dice sobre una marca o producto, seguir las opiniones o conversaciones de determinados usuarios influyentes, detectar tendencias en internet, etc.

(Joyanes, 2016) declara que el análisis de sentimientos es un método que intenta traducir las emociones humanas en datos, pero con el uso de las herramientas modernas se puede conseguir que la espontaneidad e inmediatez de la opinión en medios sociales haga que dichos sentimientos sean más auténticos y preserven su contenido emocional. El análisis relativo a contenidos no estructurados se puede medir con las siguientes características fundamentales:

1. Polaridad: positiva, negativa o incluso neutra en cuestión de una opinión.

2. Intensidad: grado de emoción que se expresa.
3. Subjetividad: la fuente que emite la expresión es objetiva, parcial o imparcial.

La minería de sentimientos puede tener distintas aplicaciones como:

- Medida de la satisfacción de los empleados y del clima laboral.
- Medida de la satisfacción del cliente.
- Prevenir abandono de clientes detectando situaciones de riesgo de pérdida de un cliente mediante la detección de opiniones negativas que se interpreten como posible abandono del cliente.
- Comparación con la competencia mediante la evaluación de la opinión sobre la competencia de marca, empresa, productos... y compararla con la nuestra.
- Detección de fortalezas y debilidades en diferentes áreas de nuestra organización, mediante la detección de opiniones positivas o negativas de impacto.
- Medida del impacto en la reputación corporativa.
- Predicción de la evolución de determinadas acciones, lanzamientos de productos, etc.
- Análisis de a opinión del electorado en el caso de votaciones políticas.

El análisis de sentimientos se encuadra dentro del procesamiento de lenguaje natural (PLN), la inteligencia artificial y de la minería de textos, entre otras técnicas, ya que fundamentalmente busca extraer información subjetiva de un texto, como un tuit, un post en un blog, etc.

AGRADECIMIENTOS

Agradezco a Dios por todas sus bendiciones, igualmente por la oportunidad de trabajar en el proceso de mejorarme a mí misma. Agradezco a mis padres por apoyarme en todo momento en esta nueva aventura, al Consejo Nacional de Ciencia y Tecnología por su apoyo en mis estudios de posgrado, al Instituto Tecnológico de Orizaba, a la Maestría de Ingeniería Administrativa, así como a la materia de Fundamentos de Ingeniería Administrativa, por proporcionarme las bases necesarias para ser mejor como profesionista y ser humano.

REFERENCIAS CONSULTADAS

- Braga, L. P. V., Valencia, L. I. O., & Carvajal, S. S. (2009). *Introducción a la Minería de Datos*. Río de Janeiro: Editora E-papers.
- Joyanes, L. A. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor.
- Matallana, F. E., & Delgado, J. M. C. (2010). *Big to small: Las estrategias de las grandes corporaciones al alcance de las medianas empresas*. Netbiblo.
- Pérez, C. L. (2007). *Minería de datos: técnicas y herramientas*. Editorial Paraninfo.
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 10(29). Recuperado a partir de <http://www.redalyc.org/resumen.oa?id=92502902>