

La información como un bien de valor incalculable

Dentro de los modelos económicos neoclásicos, abunda un supuesto que permite establecer condiciones con el fin de explicar el comportamiento de los agentes dentro del mercado. No obstante, este supuesto raramente se encuentra en situaciones reales: Este supuesto es la ***información perfecta***.

En un mundo ideal, todos los agentes que compiten dentro del mercado tienen perfecto conocimiento sobre las preferencias de sus consumidores y las decisiones de sus competidores. Además, los consumidores se encuentran perfectamente conscientes de todos los hechos que ocurren en el mercado y toman decisiones con base en esta información perfecta.

La información perfecta hace referencia a una situación en la que todos los agentes se encuentran conscientes en su totalidad de todos los hechos ocurridos en el mercado. Sin embargo, la información perfecta es una situación imposible de alcanzar debido a que los mercados se encuentran en una constante e imparable transformación.

Las empresas tienen la necesidad de generar metodologías que les permitan definir tendencias y preferencias en sus potenciales consumidores, así como también analizar los movimientos potenciales de sus competidores. Conocer todas estas variables tiene un valor incalculable para la empresa en cuestión debido a que la sobrevivencia y crecimiento de la empresa está ligada a las decisiones que se tomen.

No obstante, la información de los mercados tiene a ser lo menos estática posible. Debido a las tendencias sociales, movimientos políticos, creencias religiosas y tendencias ideológicas, los consumidores adaptan sus ideas y preferencias de forma constante. Por lo tanto, conocer las preferencias del mercado en un momento determinado no es garantía de que esta información permanezca vigente en el largo plazo.

Las empresas requieren de diseñar herramientas que les permitan obtener y analizar grandes cantidades de información con el fin de identificar tendencias que les permitan tomar la mejor elección posible. En un mundo donde existen millones de preferencias distintas, las empresas deben de seleccionar la opción que le permita satisfacer la mayor cantidad de clientes posible.

Tener la información correcta permite a los individuos la toma de decisiones adecuada a cada situación. En un escenario donde se carece de información suficiente, es claro que la tendencia tiende al fracaso o a la suerte. Debido a que la información es fundamental para la toma de decisiones, su valor es incalculable.

En el mundo real, la información perfecta es un elemento inexistente. La información tiene un costo y las empresas necesitan incluir dentro de sus costos los gastos realizados para obtener la información pertinente. Una organización que obtiene información de forma eficiente es una organización con potencial infinito.

¿Cómo pueden las empresas obtener información sobre las tendencias del mercado? Con la evolución tecnológica, los usuarios se encuentran cada vez más cerca de las empresas. Hace algunos años las empresas obtenían sus datos por medio de encuestas realizadas por ellos mismos hacia sus consumidores. No obstante, con la introducción del internet como medio de comunicación masiva, los individuos pueden mostrar sus preferencias por medio de redes sociales en las cuales se utilizan formas de identificación de las preferencias.

Las empresas pueden obtener información del mercado por diversos medios, los cuales dependiendo de su metodología y del tamaño de la muestra seleccionada tienen diferente grado de veracidad. No obstante, estos análisis normalmente resultarán en una gran base de datos en bruto, de la cual no se puede obtener una tendencia a simple vista.

Para analizar una base de datos, la cual contiene millones de datos cualitativos y cuantitativos es necesario utilizar metodologías que nos permitan identificar las partes más importantes de la información, las tendencias y las oportunidades. Existe una analogía entre el análisis de bases de datos y la minería. En la minería

se emplea una gran cantidad de recursos, no obstante, requieren de una planificación específica para ser utilizados.

Una compañía minera no gasta millones de dólares realizando una excavación sin antes realizar un análisis específico sobre el contenido del subsuelo. La empresa no se introduce al azar sobre un territorio y comienza una mina. De la misma manera, un analista de datos tiene metodologías específicas que le evitan realizar un análisis en el cual gasta recursos extra.

La minería de datos es un proceso que involucra a múltiples ciencias y disciplinas que van desde la psicología, pasando por la estadística, computación, matemáticas llegando a utilizar incluso tecnología aplicada en la inteligencia artificial. La minería de datos tiene por objetivo identificar tendencias que permitan a los analistas tomar decisiones correctas.

El uso de minería de datos no corresponde solamente al análisis de mercados, esta herramienta puede ser utilizada dentro de toda investigación que requiera el análisis de grandes cantidades de información. No obstante, dentro de este breve ensayo se analizará el uso de la herramienta dentro del contexto de los mercados y las organizaciones

¿Qué es la minería de datos?

Tomando en cuenta la analogía en la cual se hace referencia hace algunos párrafos, se puede construir un vago concepto de lo que es la minería de datos. De acuerdo a Vallejos (2006, pág. 11), la cual cita a Fayyad y a otros (1996) define que la minería de datos es **“Un proceso no trivial de identificación validad, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en datos.”**

Esta definición técnica ofrece ciertos conceptos fundamentales para comprender el uso de la herramienta. Vallejos (2006) cita también a Molina y a otros (2001) para explicar el concepto desde el punto de vista empresarial definiendo el concepto como: **“La integración de un conjunto de áreas que tienen como**

propósito la identificación de un conjunto de áreas a partir de las bases de datos que aporten un sesgo hacia la toma de decisión.”

El propósito de la minería de datos es el análisis de la información disponible que busca encontrar patrones que determinen el flujo de las acciones de las organizaciones. La minería de datos es una etapa dentro de un proceso mucho más grande, conocido como extracción del conocimiento de las bases de datos.

Las disciplinas que involucran al análisis de bases de datos involucran a la estadística, la inteligencia artificial, la computación gráfica y la capacidad de procesamiento masivo. Sería imposible pensar en una adecuada metodología de análisis de datos sin un poder computacional capaz de administrar y calcular millones de resultados por minuto.

No obstante, de acuerdo con Vallejos (2006) la idea de la minería de datos no es un concepto que surja de la mano con el nacimiento de la computación moderna. Para explicar los conceptos de la minería de datos se requiere comprender el concepto de ***“Descubrimiento de conocimiento en bases de datos”***. Este concepto es un elemento fundamental para poder realizar análisis de datos correctos fundamentados en metodologías correctas.

Descubrimiento de conocimiento en bases de datos

Con la evolución de la computación en el siglo XX, el costo de almacenamiento de la información ha decrecido de forma importante además de disminuir los costos de procesamiento de la propia información. Con la disminución de costos, el análisis de la propia información ha sido transformado hasta poder realizar estudios sumamente detallados con un costo reducido.

No obstante, de nada sirve tener una gran cantidad de datos si no se puede analizar la información oculta que forman los patrones dentro del bruto de la información. El valor real de los datos se encuentra en la información que se puede extraer de ella. Los negocios exitosos se fundamentan en la correcta

exploración de patrones y de la toma de decisiones basada en la anticipación y la preparación.

De acuerdo a Vallejos (2006) la capacidad de producir y analizar la información del mundo ha crecido tanto que esta se duplica cada 20 meses. Las organizaciones realizan secuencias SQL¹ para poder obtener la información básica. No obstante, requieren de técnicas más avanzadas para poder definir las tendencias más importantes de los datos.

El C.B.D. de acuerdo a Vallejos (2006) apunta a procesar automáticamente grandes cantidades de datos para encontrar conocimientos útiles en ellos. De esta manera el usuario puede utilizar la información para su conveniencia. El conocimiento tiene una jerarquía específica que requiere ser analizada de lo general a lo particular.

Técnicas basadas en minería de datos

Los fundamentos de la minería de datos son el resultado de un largo proceso de investigación. El desarrollo de las técnicas comenzó cuando se almaceno la información en un equipo de cómputo. La minería de daos depende de 3 tecnologías las cuales son:

- Recolección masiva de datos.
- Potentes computadoras con microprocesadores.
- Algoritmos de minería de datos.

Vallejos (2006) menciona que las bases de datos de disponibilidad comercial están creciendo a un ritmo sin precedentes. Los algoritmos de minería de datos encuentran la forma de superar conscientemente los métodos estadísticos clásicos.

¹ Las sentencias SQL son un lenguaje declarativo de acceso a bases de datos, que permiten especificar diversos tipos de operaciones. Estas sentencias permiten el manejo del algebra y el cálculo relacional con el fin de recuperar información. Fuente: (Tips And Tips, 2012)

Las principales características y objetivos de la minería de datos son los siguientes: (Vallejos, 2006)

- Explorar los datos que se encuentran en las profundidades de las bases de datos guardadas en los almacenes de datos.
- Los datos se pueden obtener de fuentes de internet o intranet.
- El entorno de la minería de datos guarda una arquitectura cliente – servidor.
- Las herramientas permiten extraer el mineral de la información enterrado en registros públicos.
- El minero es un usuario final que esta empoderado por las barrenadoras de datos.
- El hurgar datos permite encontrar resultados inesperados.
- Las herramientas de la minería de datos se combinan fácilmente y analizarse apropiadamente.
- La minería produce 5 tipos de información
 - Asociaciones
 - Secuencias
 - Clasificaciones
 - Agrupamientos
 - Pronósticos

La minería de datos guarda un método científico, debido a que formula la hipótesis en donde se diseña el experimento para coleccionar datos. Con este sistema se puede obtener un nuevo conocimiento. La minería de datos presenta un enfoque exploratorio más que confirmador.

Los alcances de la minería de datos

La tecnología de minería de datos ha tenido múltiples avances en los últimos años. Con la tecnología actual se puede generar nuevas oportunidades de negocios al proveer nuevas capacidades. No obstante, los costos de la minería de

datos tienden a aumentar con el aumento en el grado de especialización. De acuerdo a Vallejos (2006) los alcances son los siguientes:

- ***Predicción automatizada de tendencias y comportamientos:***

La minería de datos permite automatizar el proceso de encontrar información predecible en grandes bases de datos. Las preguntas normalmente requieren un análisis manual, no obstante, ahora pueden ser contestadas directamente.

Un ejemplo en donde se observa este análisis es cuando se realiza el marketing sistemático apuntando a objetivos. La minería de datos usa análisis de resultados para direccionar nuevas campañas de marketing. Con esta técnica podemos identificar el comportamiento de ciertos sectores poblacionales y repetir sus comportamientos.

- ***Descubrimiento automatizado de modelos previamente conocidos:***

Las herramientas de la minería de datos permiten identificar modelos anteriormente conocidos en un solo paso. Es posible que este método también identifique transacciones fraudulentas en sistemas bancarios y encontrar anomalías.

- ***Análisis masivo automatizado:***

Cuando las técnicas de automatización son implementadas en los sistemas de procesamiento paralelos, es posible analizar bases de datos en minutos. Los usuarios tienen la capacidad de realizar análisis automáticos en cuestión de minutos cada vez más complejos. Esta alta velocidad permite realizar mejores predicciones.

¿Cómo resolver un problema con la minería de datos?

Habiendo definido que la minería de datos es un proceso con el cual es posible detectar información en grandes conjuntos de datos, podemos aproximarnos a la metodología que utiliza la herramienta. Debemos estar conscientes de que el avance fundamental de la minería es el analizar relaciones complejas que no son visibles con técnicas convencionales.

El SQL server de Microsoft (Microsoft, 2014) pone a nuestra disposición una metodología de sencilla comprensión para los principiantes en el asunto del análisis de datos. Como mencionamos anteriormente, las secuencias SQL tienen una profunda utilidad a la hora de realizar análisis sencillos, pero se requiere de técnicas más avanzadas para obtener toda la información. En este caso analizaremos la metodología del SQL de manera introductoria.

Para realizar un modelo de minería de datos es necesario definir los siguientes momentos (Microsoft, 2014):

1. Definir el problema
2. Preparar los datos
3. Explorar los datos
4. Generar modelos
5. Explorar y validar los modelos
6. Implementar y actualizar los modelos

No obstante, este proceso no tiene una sola dirección, sino que es cíclico. Después de implementar el modelo, es necesario realizar nuevamente el proceso para confirmar que el desarrollo de nuevos modelos es posible. La minería de datos que permite realizar el SQL tiende a mejorarse a sí misma.

Definiendo el problema:

El primer paso del proceso de minería de datos, tal como se resalta en el siguiente diagrama, consiste en definir claramente el problema y considerar formas de usar los datos para proporcionar una respuesta para el mismo. (Microsoft, 2014)

Preparando los datos:

La limpieza de datos no solamente implica quitar los datos no válidos o interpolar valores que faltan, sino también buscar las correlaciones ocultas en los datos, identificar los orígenes de datos que son más precisos y determinar qué columnas son las más adecuadas para el análisis. (Microsoft, 2014)

Explorando los datos:

Al explorar los datos para conocer el problema empresarial, puede decidir si el conjunto de datos contiene datos defectuosos y, a continuación, puede inventar una estrategia para corregir los problemas u obtener una descripción más profunda de los comportamientos que son típicos de su negocio. (Microsoft, 2014)

Generando el modelo:

Antes de procesar la estructura y el modelo, un modelo de minería de datos simplemente es un contenedor que especifica las columnas que se usan para la entrada, el atributo que está prediciendo y parámetros que indican al algoritmo cómo procesar los datos. (Microsoft, 2014)

Explotando y validando el modelo:

El conjunto de datos de entrenamiento se utiliza para generar el modelo y el conjunto de datos de prueba para comprobar la precisión del modelo mediante la creación de consultas de predicción. (Microsoft, 2014)

Implementando y actualizando el modelo:

Una vez que los modelos de minería de datos se encuentran en el entorno de producción, puede llevar a cabo diferentes tareas, dependiendo de sus necesidades como el uso de modelos de predicciones, crear consultas estadísticas o crear informes. (Microsoft, 2014)

El futuro del análisis de datos

Con la evolución de las redes sociales, los individuos con el poder adquisitivo suficiente para acceder a internet se han convertido en una fuente de información infinita y de valor incalculable. Actualmente los propios usuarios externalizan sus gustos y hábitos de consumo por internet, con lo cual obtener datos es más sencillo.

Con esta clase de control dentro de los sistemas de información, las empresas tienen la capacidad de conocer los hábitos de consumo de cada individuo y generar publicidad acorde con lo demostrado en la información que el usuario pone disponible.

Actualmente la publicidad por internet se fundamenta principalmente por el estudio individual de las preferencias de los individuos. Es fundamental que el desarrollo de la minería de datos permita generar sistemas que no solo identifiquen tendencias, sino que también identifiquen comportamientos individuales.

Propuesta de tesis:

Se propone el tema: ***“Minería de datos: Herramienta de uso para el marketing orientado al consumidor individual”*** con el fin de desarrollar herramientas de minería de datos que permitan gestionar la información disponible por las redes sociales y enfocarlas a objetivos de publicidad dirigida.

Los objetivos de la tesis son:

- Desarrollo de minería de datos
- Desarrollo de técnicas de marketing
- Aplicación técnica de análisis de datos

Bibliografía

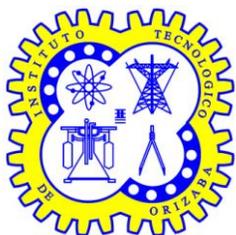
Microsoft. (2014). *SQL server 2014*. Obtenido de [https://msdn.microsoft.com/es-mx/library/ms174949\(v=sql.120\).aspx](https://msdn.microsoft.com/es-mx/library/ms174949(v=sql.120).aspx)

Tips And Tips. (2012). *Sentencias básicas en SQL*. Obtenido de <https://cwflores.wordpress.com/2008/09/11/sentencias-basicas-en-sql/>

Vallejos, S. (2006). *Universidad Nacional del Nordeste*. Obtenido de Minería de datos: http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf

Contenido

La información como un bien de valor incalculable	1
¿Qué es la minería de datos?	3
Descubrimiento de conocimiento en bases de datos	4
Técnicas basadas en minería de datos	5
Los alcances de la minería de datos	6
¿Cómo resolver un problema con la minería de datos?	8
Propuesta de tesis:	10
Bibliografía	11



Instituto Tecnológico de Orizaba

Fundamentos de Ingeniería Administrativa

Minería de datos

Profesor:

Dr. Fernando Aguirre y Hernández

Alumno:

Lic. Eusebio Alejandro Bolaños Ruiz



Orizaba Veracruz, 27 de marzo de 2016