

Application of data mining as a strategic tool in knowledge management

Diginomics.

2024; 3:123

DOI: 10.56294/digi2024123

ISSN: 3072-8428

Aplicación de la minería de datos como herramienta estratégica en la gestión del conocimiento

Chanith Uriarte del Águila¹  

¹Universidad Nacional De San Martín, Facultad de Ingeniería de Sistemas e Informática. Escuela Profesional de Ingeniería de Sistemas e Informática. Tarapoto, Perú.

Citar como: Uriarte del Águila C. Application of data mining as a strategic tool in knowledge management. Diginomics. 2024; 3:123. <https://doi.org/10.56294/digi2024123>

Autor para la correspondencia: Chanith Uriarte del Águila 

ABSTRACT

Introduction: the study analyzed the importance of data mining as a fundamental tool for transforming large volumes of information into knowledge applicable to business and academic decision making. This discipline allowed the discovery of hidden patterns and significant relationships between variables, facilitating the prediction of behaviors and the optimization of strategic processes.

Development: during the research, several studies were reviewed that evidenced the effectiveness of data mining techniques in different contexts. A study applied decision trees and classification methods in Argentinean institutions, managing to identify determining factors in student dropout. A study integrated fuzzy logic and data mining to predict school dropout in Mexico, demonstrating the usefulness of the hybrid approach. Likewise, a study highlighted its application in the business environment, where data mining made it possible to segment customers, personalize strategies and improve organizational competitiveness. In the particular case of Telefónica del Perú, tools such as Weka, SQL Server and SPSS were used, confirming that technological integration strengthens management and analysis processes.

Conclusions: the research concluded that data mining significantly improved decision making in the customer management area of Telefónica del Perú, Tarapoto zone. It was proven that its implementation allowed automating the analysis, reducing errors and generating accurate information for strategic decisions. Consequently, the hypothesis that data mining tools have a positive impact on organizational efficiency and foster a culture based on knowledge and innovation was validated.

Keywords: Data Mining; Decision Making; Academic Performance; B2B Marketing; Organizational Intelligence.

RESUMEN

Introducción: el estudio analizó la importancia de la minería de datos como herramienta fundamental para transformar grandes volúmenes de información en conocimiento aplicable a la toma de decisiones empresariales y académicas. Esta disciplina permitió descubrir patrones ocultos y relaciones significativas entre variables, facilitando la predicción de comportamientos y la optimización de procesos estratégicos.

Desarrollo: durante la investigación, se revisaron diversos estudios que evidenciaron la eficacia de las técnicas de minería de datos en distintos contextos. Estudios aplicaron árboles de decisión y métodos de clasificación en instituciones argentinas, logrando identificar factores determinantes en la deserción estudiantil. Un estudio integró lógica difusa y minería de datos para predecir el abandono escolar en México, demostrando la utilidad del enfoque híbrido. Asimismo, un estudio destacó su aplicación en el ámbito empresarial, donde la minería de datos permitió segmentar clientes, personalizar estrategias y mejorar la competitividad organizacional. En el caso particular de Telefónica del Perú, se emplearon herramientas como Weka, SQL Server y SPSS, confirmando que la integración tecnológica fortalece los procesos de gestión y análisis.

Conclusiones: la investigación concluyó que la minería de datos mejoró significativamente la toma de decisiones en el área de gestión al cliente de Telefónica del Perú, zonal Tarapoto. Se comprobó que su implementación permitió automatizar el análisis, reducir errores y generar información precisa para decisiones estratégicas. En consecuencia, se validó la hipótesis de que las herramientas de minería de datos inciden positivamente en la eficiencia organizacional y fomentan una cultura basada en el conocimiento y la innovación.

Palabras clave: Minería de Datos; Toma de Decisiones; Rendimiento Académico; Marketing B2B; Inteligencia Organizacional.

INTRODUCCIÓN

En la actualidad, el uso de la minería de datos se ha consolidado como una herramienta clave para transformar grandes volúmenes de información en conocimiento útil para la toma de decisiones empresariales y académicas. Este enfoque, basado en la aplicación de técnicas estadísticas, algoritmos y métodos computacionales, permite descubrir patrones ocultos en los datos, facilitando la predicción de comportamientos, la identificación de tendencias y la optimización de procesos estratégicos. Su relevancia se extiende a múltiples áreas, entre ellas la educación, el marketing y la gestión organizacional, donde se emplea para analizar el rendimiento académico, predecir la deserción estudiantil, comprender el comportamiento del consumidor o mejorar la eficiencia operativa.

Diversos estudios han demostrado la aplicabilidad y eficacia de la minería de datos en distintos contextos. Sposito⁽¹⁾ implementó árboles de decisión y algoritmos de clasificación para evaluar el rendimiento y la deserción estudiantil en la Universidad Nacional de La Matanza, evidenciando la capacidad predictiva de estas técnicas sobre variables académicas y socioeconómicas. De forma similar, un estudio aplicó metodologías supervisadas y no supervisadas en la Universidad Nacional de Misiones, destacando la precisión de los modelos de clasificación frente a los de agrupamiento, y subrayando la necesidad de considerar factores sociales y demográficos en el análisis del abandono universitario. Por su parte, Domínguez⁽²⁾ incorporó lógica difusa a la minería de datos para desarrollar un sistema predictivo del abandono escolar en instituciones privadas de México, demostrando la eficacia del enfoque híbrido entre algoritmos y razonamiento humano.

En el ámbito empresarial, Altamiranda et al.⁽³⁾ resaltaron que la minería de datos se ha convertido en una herramienta estratégica para la competitividad y la toma de decisiones en marketing, especialmente en entornos B2B, donde facilita la segmentación de clientes, la personalización de estrategias y la optimización de recursos. De este modo, la minería de datos no solo potencia la capacidad analítica de las organizaciones, sino que también promueve una gestión basada en la evidencia, la automatización y la innovación.

En síntesis, la creciente generación de datos en la era digital ha impulsado la adopción de técnicas de minería de datos como soporte esencial para la toma de decisiones informadas. Este proceso, al integrar conocimiento, tecnología y análisis inteligente, permite a las instituciones —tanto educativas como empresariales— avanzar hacia una gestión más eficiente, predictiva y sustentada en el valor de la información.

DESARROLLO

Antecedentes de la Investigación

Sposito⁽¹⁾ “Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil”. La evaluación del rendimiento académico y de la deserción estudiantil de los estudiantes del Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT) de la Universidad Nacional de La Matanza (UNLAM). La investigación se realizó aplicando el proceso de descubrimiento de conocimiento sobre los datos de alumnos del periodo 2003 – 2008.

En el presente proyecto se tomaron en cuenta las siguientes variables:

- Datos del estudiante.
- Datos de las carreras del DIIT.

- Datos de los planes de estudio, vigentes y no vigentes, de las carreras.
- Datos de las materias de los planes de estudio.
- Datos de las notas, por carrera, plan de estudio de materia, de los estudiantes.
- Datos de los censos realizados a estudiantes.

Técnicas de Minería de Datos Aplicada

Se utilizaron los arboles de decisión implementando la tarea de clasificación bajo el algoritmo J48 (implementación en Weka del algoritmo C4.5) y el FT como algoritmo de minería.

Herramientas utilizadas: Ms SQL Server para la generación de un almacén de datos, el software SPSS para realizar un pre procesamiento de los datos y el software Weka (Waikato Environment for knowledge Analysis) para encontrar un clasificador de rendimiento académico y para detectar los patrones determinantes de la deserción estudiantil.

El estudio “Minería de Datos aplicada al análisis de la deserción en la carrera de Analista en sistemas de computación”. Tesis de grado para la licenciatura en sistemas de información. En esta investigación se realizó Minería de Datos sobre el Cubo 04 Desgranamiento, exportado del Sistema de Gestión Académica

SIU-Guaraní, provistos por el Ministerio de Educación, Ciencia y Tecnología de la Nación. Se ha experimentado con métodos Supervisados, como la Clasificación y No Supervisados como es el Agrupamiento. El objetivo principal fue maximizar la calidad que los modelos tienen para clasificar y agrupar a los estudiantes, de acuerdo a sus características académicas, factores sociales y demográficos, que han desertado de la Carrera Analista en Sistemas de Computación de la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones. El proyecto se desarrolló bajo la metodología de libre difusión Crisp-DM y con la herramienta comercial IBM DB2 Warehouse (versión 9.5). Los resultados obtenidos permitieron observar, no sólo los diferentes patrones de los alumnos desertores, sino que también las características de los estudiantes en actividad y egresados. La calidad de los modelos obtenidos a través de la clasificación con árboles de decisión superó a la técnica de agrupamiento a través de la generación de clústeres y ambas han superado ampliamente lo planteado. Si bien los expertos en el estudio de la deserción universitaria han elogiado la capacidad que los modelos tienen para distinguir a cada tipo de alumno, han planteado la necesidad de contemplar más variables socio económicas en el estudio de la deserción.

Domínguez⁽²⁾. “Minería de Datos y lógica difusa como método para la predicción del abandono escolar de alumnos de institutos de nivel superior privado”. Sistema de predicción para la detección de factores que influyen para el abandono escolar de alumnos que estudian en instituciones privadas de nivel superior. Se utilizó la minería de datos (CRISP-DM) y la lógica finalmente, el sistema de predicción se aplicó a una institución difusa, como técnicas de análisis. El sistema se aplicó en el instituto de universitaria privada en un periodo comprendiendo un ciclo Estudios Superiores del Centro de Chiapas, México para corroborar su funcionamiento.

El objetivo fue desarrollar un sistema de predicción utilizando lógica difusa con la metodología de desarrollo de minería de datos para predecir el abandono escolar de alumnos de institutos de nivel superior privado.

Técnicas de Minería de Datos Aplicada: reglas Difusas con la tarea de Clasificación

Herramientas utilizadas: el sistema de inferencia difuso fue construido en Matlab. Altamiranda et al.⁽³⁾. “Minería de datos como herramienta para el desarrollo de estrategias de mercadeo B2B en sectores productivos, afines a los colombianos: una revisión de casos”, en Sotavento MBA, N° 22, pp. 126-136. Actualmente se estima que el suministro de datos del mundo se duplica cada 20 meses. Lo anterior implica un crecimiento excesivo en el volumen de datos que se maneja tanto en la comunidad científica como en los sectores productivos de la economía, que ha sobrepasado la capacidad humana de analizar, resumir y extraer conocimientos a tales cantidades de datos. Lo anterior, hace necesaria una nueva generación de herramientas capaces de automatizar el análisis de los datos almacenados. El conjunto de estas herramientas lo estudia un nuevo campo de investigación llamado minería de datos.

La minería de datos se ha convertido en una herramienta estratégica para la toma de decisiones de mercadeo, producción, organización y demás factores en la empresa, que de cierta manera la hacen más competitiva. El presente artículo analiza la manera como la minería de datos, una técnica ampliamente relacionada con la investigación de operaciones, incide en el diseño de estrategias de mercadeo B2B, en empresas de sectores industriales. Esta investigación guarda relación directa con proyectos investigativos realizados por la Universidad Autónoma del Caribe, y relacionados con el desarrollo de la competitividad e innovación de los sectores estratégicos del departamento del Atlántico.

Definición de términos

Toma de Decisiones

Según el libro “Toma de decisiones gerenciales: métodos cuantitativos para la administración” menciona que la toma de decisiones empresariales constituye un proceso de selección entre las diferentes alternativas posibles.

A partir de la elección de una alternativa se podrá pasar al proceso de la planificación de actividades, desarrollo de las estrategias y formulación de programas y presupuestos. Para dirigir eficazmente la pequeña y mediana empresa se necesita utilizar los recursos disponibles de una manera eficiente, al ser los recursos limitados los gerentes deberán tratar de aprovechar todas las oportunidades.

La toma de decisiones se debe realizar en los diferentes niveles de la empresa, esto trae consigo que las decisiones individuales, correspondientes a cada nivel de responsabilidad, afectan de una manera directa al desarrollo de todo el sistema organizativo y condicionan tanto la consecución de los objetivos particulares como la de los objetivos generales o globales. En la siguiente figura se observa las fases de la planificación de toma de decisiones en función del análisis del valor.

Minería de datos

El gran desarrollo tecnológico de los últimos años ha potenciado el almacenamiento de grandes volúmenes de información. No solo desde el punto de vista del “computador personal” sino también desde el punto de vista de la computación como un conjunto de dispositivos electrónicamente interconectados capaces de soportar el transporte de grandes cantidades de información en una dirección determinada con el propósito de ser almacenada centralizada o descentralizada,

o bien desde el punto de vista de redes electrónicas donde desde cada nodo se permite la posibilidad de cargar información en una o más bases de datos.

Al mayor poder de procesamiento y transporte de información se suman los avances tecnológicos en materia de base de datos que soportan “grandes” volúmenes de información de diferentes orígenes y fuentes pudiendo incluir texto, números, imágenes, video, etc.

Estos avances en materia de gran capacidad de almacenamiento de información han dado lugar a la creación de nuevas tecnologías como la Minería de Datos.

Conceptualmente la Minería de Datos o Explotación de Datos, como también será llamada en el presente trabajo, se puede definir como un conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos con motivo de predecir de forma automatizada tendencias y comportamientos y/o descubrir de forma automatizada modelos previamente desconocidos.

Algoritmos 4g

Las comunicaciones móviles se han convertido, en los últimos años, en uno de los focos principales de las economías nacionales, tanto desde el punto de vista de ingresos generados como inversiones realizadas por los diferentes operadores. Esta creciente importancia se debe a un cambio en la demanda de los servicios y aplicaciones por parte del usuario, empleando sus terminales móviles (tablets, smartphones) como principal punto de acceso a Internet.

Este acceso se realiza a través del servicio de banda ancha móvil MBAS. Para la provisión de este servicio los operadores pueden utilizar tecnologías avanzadas de 3G como High Speed Packet Access (HSPA) así como las futuras 4G, Long Term Evolution (LTE). HSPA se está implementando actualmente en España, LTE está previsto para el 2015, pero, en el actual panorama de crisis, es de vital importancia para los operadores maximizar el rendimiento de su inversión en equipamiento de red.

Para ello se deben realizar estudios comparativos del coste de inversión en la red celular de acceso de ambas tecnologías para diferentes escenarios de tráfico de usuario a nivel nacional. Esto implica la utilización de herramientas software de planificación de redes que incorporen algoritmos precisos de planificación de ambas tecnologías.

Olap

Proporciona un modelo de datos intuitivo y conceptual, para que los usuarios que no tengan experiencia como analistas puedan comprender y relacionar los datos mostrados. Este modelo es llamado análisis multidimensional, siendo habilitado para ver los datos a través de múltiples filtros, o dimensiones.

Los sistemas OLAP organizan los datos directamente como estructuras multidimensionales, incluyendo herramientas fáciles de usar por usuarios para conseguir la información en múltiples y simultáneas vistas dimensionales. OLAP es también rápido para el usuario.

Rápidos tiempos de respuesta permiten que los gerentes y analistas puedan preguntar y resolver más situaciones en un corto período de tiempo. Una dimensión es una vista de los datos categóricamente consistente. Una característica de

las dimensiones es la habilidad de hacer slice-and-dice. Slice (rebanada) y dice (cubo) hacen particiones de los datos en una base de datos multidimensional de acuerdo a los valores de ciertas dimensiones.

Otra capacidad inherente en el diseño de OLAP es la rotación y anidamiento (Pivoting-and-Nesting) de las dimensiones. El pivoted permite rotar los datos desde las columnas hasta las filas. También es importante mencionar el concepto de drill, que en los sistemas OLAP tiene un significado muy específico. Drill down es la acción de seleccionar un miembro para ver el siguiente nivel inferior de detalle en la jerarquía. Drill up es seleccionar un miembro para ver el siguiente nivel superior, esto es, una acción de bottom-up. La mecánica o funcionamiento de las interfaces OLAP, especialmente pointing-and-clicking (apuntar y seleccionar) para hacer drill-down dentro de las capas de interés se hace posible por la velocidad con que las consultas son resueltas.

Esta funcionalidad permite por completo a los gerentes y analistas un nuevo proceso para tratar con grandes cantidades de datos, un proceso conocido con el nombre de análisis ad hoc. En resumen, los sistemas OLAP organizan los datos por intersecciones multidimensionales. Esta organización, acompañada por una herramienta de interface para rotar y anidar dimensiones, permite a los usuarios visualizar rápidamente valores en detalle, patrones, variaciones y anomalías en los datos que estarían de otra manera ocultos por un análisis dimensional simple. A mayor número de dimensiones (dentro de los límites razonables), mayor es la profundidad del análisis.

Bases teóricas

Minería de datos

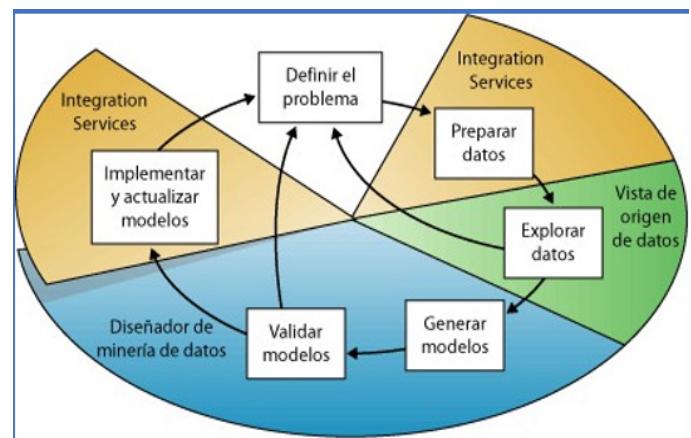
La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos.

Estos patrones y tendencias se pueden recopilar y definir como un modelo de minería de datos. Los modelos de minería de datos se pueden aplicar en escenarios como los siguientes:

- Pronóstico: cálculo de las ventas y predicción de las cargas del servidor o del tiempo de inactividad del servidor.
- Riesgo y probabilidad: elección de los mejores clientes para la distribución de correo directo, determinación del punto de equilibrio probable para los escenarios de riesgo, y asignación de probabilidades a diagnósticos y otros resultados.
- Recomendaciones: determinación de los productos que se pueden vender juntos y generación de recomendaciones.
- Búsqueda de secuencias: análisis de los artículos que los clientes han introducido en el carrito de la compra y predicción de posibles eventos.
- Agrupación: distribución de clientes o eventos en grupos de elementos relacionados, y análisis y predicción de afinidades.

El siguiente diagrama describe las relaciones existentes entre cada paso del proceso y las tecnologías de Microsoft SQL Server

que se pueden usar para completar cada paso.



Fuente: <https://msdn.microsoft.com>

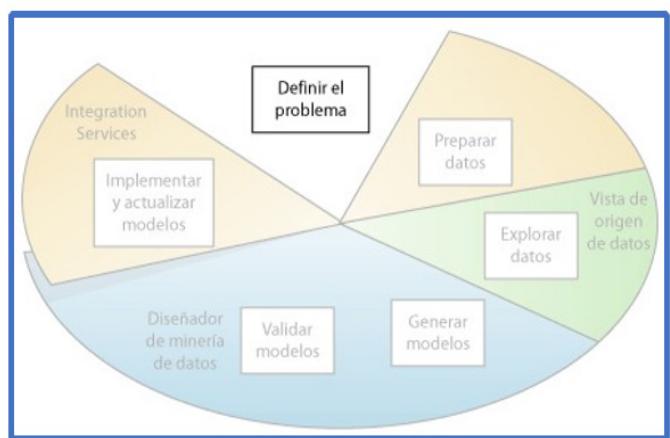
Figura 1. Interrelación de Pasos de la minería de Datos

El proceso que se ilustra en el diagrama es cíclico, lo que significa que la creación de un modelo de minería de datos es un proceso dinámico e iterativo. Una vez que ha explorado los datos, puede que descubra que resultan insuficientes para crear los modelos de minería de datos adecuados y que, por tanto, debe buscar más datos.

O bien, puede generar varios modelos y descubrir entonces que no responden adecuadamente al problema planteado cuando los definió y que, por tanto, debe volver a definir el problema.

Es posible que deba actualizar los modelos una vez implementados debido a que haya más datos disponibles. Puede que haya que repetir cada paso del proceso muchas veces para crear un modelo adecuado.

Definir 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.

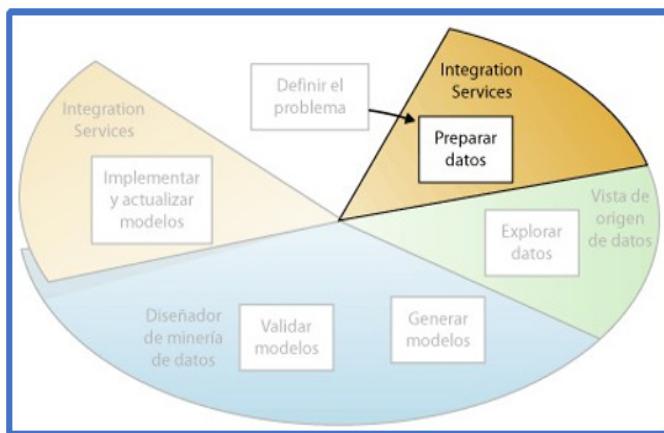


Fuente: <https://msdn.microsoft.com>

Figura 2. Primer Paso de la minería de Datos; Definir el Problema

Este paso incluye analizar los requisitos empresariales, definir el ámbito del problema, definir las métricas por las que se evaluará el modelo y definir los objetivos concretos del proyecto de minería de datos. Estas tareas se traducen en preguntas como las siguientes: ¿Qué está buscando? ¿Qué tipos de relaciones intenta buscar? ¿Refleja el problema que está intentando resolver las directivas o procesos de la empresa? ¿Desea realizar

predicciones a partir del modelo de minería de datos o solamente buscar asociaciones y patrones interesantes? ¿Qué resultado o atributo desea predecir? ¿Qué tipo de datos tiene y qué tipo de información hay en cada columna? En caso de que haya varias tablas, ¿cómo se relacionan? ¿Necesita limpiar, agregar o procesar los datos antes de poder usarlos? ¿Cómo se distribuyen los datos? ¿Los datos son estacionales? ¿Los datos representan con precisión los procesos de la empresa?



Fuente: <https://msdn.microsoft.com>

Figura 3. Segundo Paso de la minería de Datos; Preparar Datos

Para responder a estas preguntas, puede que deba dirigir un estudio de disponibilidad de datos para investigar las necesidades de los usuarios de la empresa con respecto a los datos disponibles. Si los datos no abarcan las necesidades de los usuarios, podría tener que volver a definir el proyecto.

También debe considerar las maneras en las que los resultados del modelo se pueden incorporar en los indicadores de rendimiento clave (KPI) que se utilizan para medir el progreso comercial.

Preparar los datos: el segundo paso del proceso de minería de datos, como se indica en el siguiente diagrama, consiste en consolidar y limpiar los datos identificados en el paso Definir el problema.

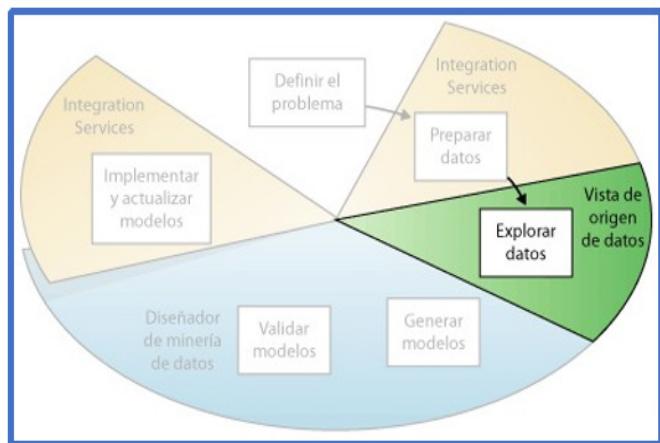
Los datos pueden estar dispersos en la empresa y almacenados en formatos distintos; también pueden contener incoherencias como entradas que faltan o incorrectas. Por ejemplo, los datos pueden mostrar que un cliente adquirió un producto incluso antes que se ofreciera en el mercado o que el cliente compra regularmente en una tienda situada a 2000 kilómetros de su casa.

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. Por ejemplo, ¿debería utilizar la fecha de envío o la fecha de pedido? ¿Qué influye más en las ventas: la cantidad, el precio total o un precio con descuento? Los datos incompletos, los datos incorrectos y las entradas que parecen independientes, pero que de hecho están estrechamente correlacionadas, pueden influir en los resultados del modelo de maneras que no espera.

Es importante tener en cuenta que los datos que se usan para la minería de datos no necesitan almacenarse en un cubo de procesamiento analítico en línea (OLAP), ni siquiera en una base de datos relacional, aunque puede usar ambos como orígenes de datos. Puede realizar minería de datos mediante cualquier origen

de datos definido como origen de datos de Analysis Services. Por ejemplo, archivos de texto, libros de Excel o datos de otros proveedores externos. Para más información, consulte Tipos de orígenes de datos admitidos (SSAS multidimensional).

Explorar los datos: el tercer paso del proceso de minería de datos, como se resalta en el siguiente diagrama, consiste en explorar los datos preparados.



Fuente: <https://msdn.microsoft.com>

Figura 4. Tercer Paso de la minería de Datos; Explorar Datos

Debe conocer los datos para tomar las decisiones adecuadas al crear los modelos de minería de datos. Entre las técnicas de exploración se incluyen calcular los valores mínimos y máximos, calcular la media y las desviaciones estándar, y examinar la distribución de los datos.

Por ejemplo, al revisar el máximo, el mínimo y los valores de la media se podrían determinar que los datos no son representativos de los clientes o procesos de negocio, y que por consiguiente debe obtener más datos equilibrados o revisar las suposiciones que son la base de sus expectativas.

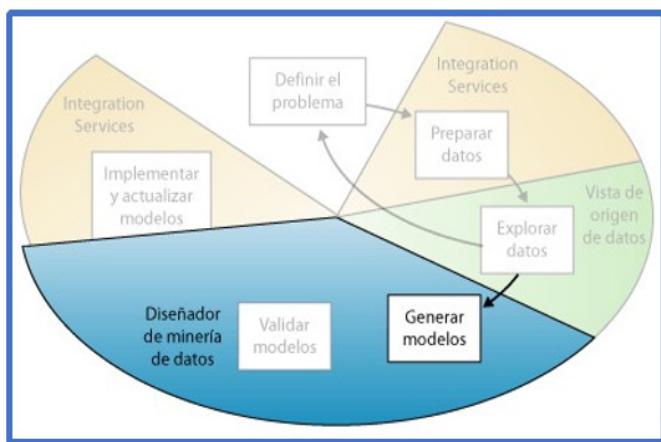
Las desviaciones estándar y otros valores de distribución pueden proporcionar información útil sobre la estabilidad y exactitud de los resultados. Una desviación estándar grande puede indicar que agregar más datos podría ayudarle a mejorar el modelo. Los datos que se desvían mucho de una distribución estándar se podrían sesgar o podrían representar una imagen precisa de un problema de la vida real, pero dificultar el ajustar un modelo a 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.

Generar modelos: el cuarto paso del proceso de minería de datos, como se resalta en el siguiente diagrama, consiste en generar el modelo o modelos de minería de datos. Usará los conocimientos adquiridos en el paso Explorar los datos para definir y crear los modelos.

Deberá definir qué columnas de datos desea que se usen; para ello, creará una estructura de minería de datos. La estructura de minería de datos se vincula al origen de datos, pero en realidad no contiene ningún dato hasta que se procesa. Al procesar la estructura de minería de datos, Analysis Services genera agregados y otra información estadística que se puede usar para el análisis. Cualquier modelo de minería de datos que esté basado en la estructura puede utilizar esta información.

Para obtener más información acerca de cómo se relacionan las estructuras de minería de datos con los modelos de minería de datos, vea Arquitectura lógica (Analysis Services - Minería de datos).



Fuente: <https://msdn.microsoft.com>

Figura 5. Cuarto Paso de la minería de Datos; Generar Modelos

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. El procesamiento de un modelo a menudo se denomina entrenamiento. El entrenamiento hace referencia al proceso de aplicar un algoritmo matemático concreto a los datos de la estructura para extraer patrones.

Los patrones que encuentre en el proceso de entrenamiento dependerán de la selección de los datos de entrenamiento, el algoritmo que elija y cómo se haya configurado el algoritmo. SQL Server 2014 contiene muchos algoritmos diferentes. Cada uno está preparado para un tipo diferente de tarea y crea un tipo distinto de modelo. Para obtener una lista de los algoritmos proporcionados en SQL Server 2014, vea Algoritmos de minería de datos (Analysis Services: Minería de datos).

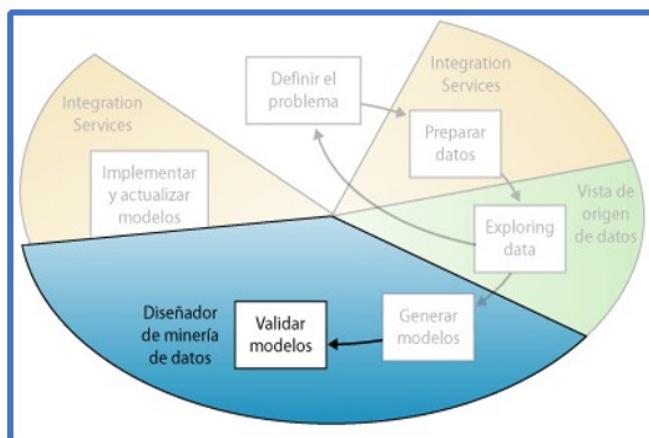
También puede utilizar los parámetros para ajustar cada algoritmo y puede aplicar filtros a los datos de entrenamiento para utilizar un subconjunto de los datos, creando resultados diferentes. Después de pasar los datos a través del modelo, el objeto de modelo de minería de datos contiene los resúmenes y modelos que se pueden consultar o utilizar para la predicción.

Puede definir un modelo nuevo mediante el Asistente para minería de datos de Herramientas de datos de SQL Server o con el lenguaje DMX (Extensión de minería de datos). Para obtener más información sobre cómo utilizar el Asistente para minería de datos, vea Asistente para minería de datos (Analysis Services - Minería de datos). Para obtener más información sobre cómo utilizar DMX, vea Referencia de Extensiones de minería de datos (DMX).

Es importante recordar que siempre que los datos cambian, debe actualizar la estructura y el modelo de minería de datos. Al actualizar una estructura de minería de datos volviéndola a procesar, Analysis Services recupera los datos del origen, incluido cualquier dato nuevo si el origen se actualiza dinámicamente, y vuelve a llenar la estructura de minería de datos. Si tiene modelos que están basados en la estructura, puede elegir actualizar estos, lo que significa que se vuelven a entrenar

con los nuevos datos, o pueden dejar los modelos tal cual. Para más información, consulte Requisitos y consideraciones de procesamiento (minería de datos).

Explorar y validar los modelos: el quinto paso del proceso de minería de datos, como se resalta en el siguiente diagrama, consiste en explorar los modelos de minería de datos que ha generado y comprobar su eficacia.



Fuente: <https://msdn.microsoft.com>

Figura 6. Quinto Paso de la minería de Datos; Validar Modelos

Antes de implementar un modelo en un entorno de producción, es aconsejable probar si funciona correctamente. Además, al generar un modelo, normalmente se crean varios con configuraciones diferentes y se prueban todos para ver cuál ofrece los resultados mejores para su problema y sus datos.

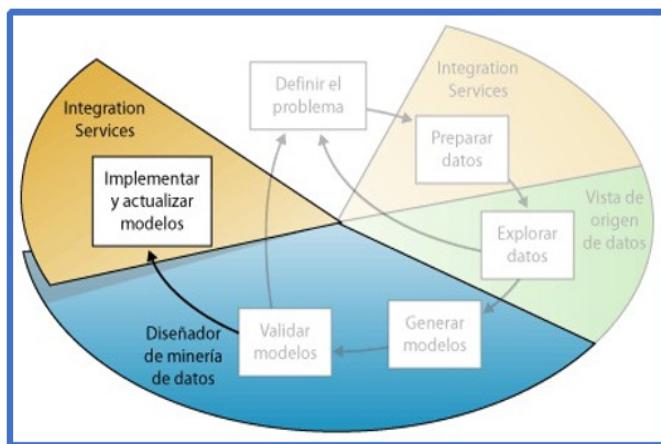
Analysis Services proporciona herramientas que ayudan a separar los datos en conjuntos de datos de entrenamiento y pruebas, para que pueda evaluar con precisión el rendimiento de todos los modelos en los mismos datos. El conjunto de datos de entrenamiento se utiliza para generar el modelo y el conjunto de datos de prueba para comprobar la precisión del modelo mediante la creación de consultas de predicción. En SQL Server 2014 Analysis Services (SSAS), estas particiones se pueden hacer automáticamente mientras se genera el modelo de minería de datos. Para más información, consulte Prueba y validación (minería de datos).

Puede explorar las tendencias y patrones que los algoritmos detectan mediante los visores del diseñador de minería de datos de Herramientas de datos de SQL Server. Para más información, consulte Visores de modelos de minería de datos. También puede comprobar si los modelos crean predicciones correctamente mediante herramientas del diseñador como el gráfico de mejora respecto al modelo predictivo y la matriz de clasificación. Para comprobar si el modelo es específico de sus datos o se puede utilizar para realizar inferencias en la población general, puede utilizar la técnica estadística denominada validación cruzada para crear automáticamente subconjuntos de los datos y probar el modelo con cada uno. Para más información, consulte Prueba y validación (minería de datos).

Si ninguno de los modelos que ha creado en el paso Generar modelos funciona correctamente, puede que deba volver a un paso anterior del proceso y volver a definir el problema o volver a investigar los datos del conjunto de datos original.

Implementar y actualizar los modelos: el último paso del proceso de minería de datos, como se resalta en el siguiente

diagrama, consiste en implementar los modelos que funcionan mejor en un entorno de producción.



Fuente: <https://msdn.microsoft.com>

Figura 7. Sexto Paso de la minería de Datos; Implementar y Actualizar Modelos

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. Las siguientes son algunas de las tareas que puede realizar:

Use los modelos para crear predicciones que luego podrá usar para tomar decisiones comerciales. SQL Server pone a su disposición el lenguaje DMX, que podrá usar para crear consultas de predicción, y el Generador de consultas de predicción, que le ayudará a generar las consultas. Para más información, consulte Referencia de Extensiones de minería de datos (DMX).

Crear consultas de contenido para recuperar estadísticas, reglas o fórmulas del modelo. Para más información, consulte Consultas de minería de datos.

Incrustar la funcionalidad de minería de datos directamente en una aplicación. Puede incluir Objetos de administración de análisis (AMO), que contiene un conjunto de objetos que la aplicación pueda utilizar para crear, cambiar, procesar y eliminar estructuras y modelos de minería de datos. También puede enviar mensajes XML for Analysis (XMLA) directamente a una instancia de Analysis Services.

Actualizar los modelos después de la revisión y análisis. Cualquier actualización requiere que vuelve a procesar los modelos. Para más información, consulte Procesar objetos de minería de datos.

Actualizar dinámicamente los modelos, cuando entran más datos en la organización, y realizar modificaciones constantes para mejorar la efectividad de la solución debería ser parte de la estrategia de implementación. Para obtener más información, consulte Administración de las soluciones y los objetos de minería de datos.

Datos, información y conocimiento

Datos

Los datos son en esencia números o texto que puede ser procesado en una computadora, en la actualidad las organizaciones acumulan grandes cantidades de datos en distintos formatos y en distintas bases de datos, entre las que se incluyen datos operacionales o transaccionales en las que se almacenan costos, ventas, inventarios, contabilidad, etc.⁽⁴⁾

Información

Los patrones, asociaciones o relaciones entre los datos proporcionan información, por ejemplo el análisis de transacciones de un punto de venta nos pueden dar información sobre qué cantidad de productos se han vendido y durante cuánto tiempo.⁽⁴⁾

Conocimiento

La información puede ser convertida en conocimiento partiendo de patrones históricos.⁽⁴⁾

Predicción (Forecasting)

Una vez descubiertas reglas importantes, estas pueden ser utilizadas para estimar algunas variables de salida. Puede ser en el caso de secuencias en el tiempo, o bien en la identificación e interrupción a tiempo, de una futura mala experiencia de crédito. En esta tarea, se complementan las técnicas estadísticas tradicionales con aquellas provenientes de la inteligencia artificial. Conceptos adaptativos como los algoritmos genéticos y las redes neuronales, permiten realizar predicciones más acertadas, especialmente en casos de gran complejidad y con relaciones internas.

Interpretación y evaluación

Finalmente se procede a su validación, comprobando que las conclusiones son válidas y satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, se alterará alguno de los procesos anteriores en busca de nuevos modelos.

Hipótesis

Hipótesis Alterna

Con la aplicación de la herramienta de MINERIA DE DATOS se mejorará el proceso de toma decisiones en el área de gestión al cliente de telefónica del Perú zonal Tarapoto.⁽⁵⁾

Hipótesis Nula

Con la aplicación de la herramienta de MINERIA DE DATOS no se mejorará el proceso de toma decisiones en el área de gestión al cliente de telefónica del Perú zonal Tarapato.

Sistema de variables

- Variable Independiente ===== Minería de Datos.
- Variable Dependiente ===== Toma de Decisiones.

VARIABLES	DIMENSIONES	INDICADORES
DEPENDIENTE		• Nivel de satisfacción del usuario.
Toma de decisiones en el área de atención al cliente	• Eficiencia	• Tiempo de asignación del objetivo
INDEPENDIENTE	• Eficiencia	• Nivel de satisfacción del usuario. • Tiempo de asignación del objetivo
Sistema de información basado en Minera Datos		

Figura 8. Cuadro de indicadores según variables

Objetivos

General

Determinar el efecto del uso de minería de datos en la toma de decisiones en el área de gestión al cliente de telefónica del Perú zonal Tarapoto.

Especifico

- Analizar el proceso de toma de decisiones en el área de gestión al cliente.
- Diseñar e Implementar el sistema con herramientas de minería de datos.
- Medir los resultados en el proceso de toma de decisiones con el uso del aplicativo con minería de datos.

CONCLUSIONES

El estudio permitió demostrar que la minería de datos constituye una herramienta esencial para transformar grandes volúmenes de información en conocimiento útil, aplicable tanto en entornos académicos como empresariales. A través del análisis de los antecedentes revisados, se evidenció que las técnicas de minería de datos, especialmente los árboles de decisión, la lógica difusa y los métodos de clasificación, ofrecen resultados altamente precisos en la identificación de patrones y en la predicción de comportamientos. Estudios confirmaron la capacidad predictiva de estas técnicas en el ámbito educativo, particularmente para comprender los factores que inciden en el rendimiento académico y en la deserción estudiantil. Asimismo, se destacó que la incorporación de variables socioeconómicas y demográficas mejora sustancialmente la precisión de los modelos, lo que resalta la importancia de integrar la dimensión humana en los procesos de análisis automatizado.

En el contexto empresarial, estudios demostraron que la minería de datos impulsa la competitividad organizacional al optimizar la toma de decisiones estratégicas, especialmente en el marketing B2B. Las empresas que aplican estas herramientas

logran identificar tendencias de consumo, segmentar eficazmente a sus clientes y personalizar sus estrategias, lo que se traduce en un uso más eficiente de los recursos y una ventaja competitiva sostenible. En consecuencia, la minería de datos no solo actúa como un instrumento técnico, sino también como un elemento estratégico que transforma la gestión empresarial hacia un modelo basado en la evidencia y el conocimiento.

De manera general, la investigación concluyó que el uso de minería de datos incide positivamente en la toma de decisiones en el área de gestión al cliente, al permitir la automatización del análisis, la reducción de errores y la obtención de información precisa y oportuna. Su implementación fomenta una cultura organizacional orientada al aprendizaje continuo y a la innovación tecnológica. Por tanto, la hipótesis alterna se confirma: la aplicación de herramientas de minería de datos mejora significativamente los procesos de toma de decisiones en Telefónica del Perú, zonal Tarapoto. En síntesis, la minería de datos representa una convergencia efectiva entre tecnología, análisis y gestión del conocimiento, consolidándose como un pilar indispensable para la eficiencia operativa, la planificación estratégica y la sostenibilidad institucional en la era digital.

FINANCIACIÓN

Ninguna.

CONFLICTO DE INTERESES

Los autores declaran que no existe conflicto de intereses.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Chanith Uriarte del Águila.

Curación de datos: Chanith Uriarte del Águila.

Ánalisis formal: Chanith Uriarte del Águila.

Redacción – borrador original: Chanith Uriarte del Águila.

Redacción – revisión y edición: Chanith Uriarte del Águila.

REFERENCIAS BIBLIOGRÁFICAS

1. Sposito O. Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil. 2008. http://www.iiis.org/CDs2010/CD2010CSC/CISCI_2010/PapersPdf/CA156FK.pdf
2. Domínguez M. Minería de datos y lógica difusa como método para la predicción del abandono escolar de alumnos de institutos de nivel superior privado. Chiapas (MX): Estudios Superiores del Centro de Chiapas; 2008. <http://pcti.mx/tesis-de-posgradoenmexico?task=callelement&format=raw&itemid=382&element=5832706c-3ae3-408b-93e3-bca7418d0376&method=download>
3. Altamiranda L, Peña A, Ospino M, Volpe I, Ortega D, Cantillo E. Minería de datos como herramienta para el desarrollo de estrategias de mercadeo B2B en sectores productivos afines a los colombianos: una revisión de casos. Sotavento MBA. 2013;22:126-136. <http://revistas.uexternado.edu.co/index.php/sotavento/article/view/3709/3978>
4. Rodrigo C. Minería de datos, una herramienta para la toma de decisiones. 2006. http://www.academia.edu/8927740/MINERIA_DE_DATOS
5. Telefónica. Telefónica refuerza sus capacidades de big data con la integración de Synergic Partners. Madrid (ES): Telefónica; 2015 jul 5. <https://www.telefonica.com/es/web/sala-de-prensa/-/telefonica-refuerza-sus-capacidades-de-big-data-con-la-integracion-de-synergic-partners>