

Minería de datos. Qué es y por qué es importante

La minería de datos es el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados. Empleando una amplia variedad de técnicas, puede utilizar esta información para incrementar sus ingresos, recortar costos, mejorar sus relaciones con clientes, reducir riesgos y más. El proceso de hurgar en los datos para descubrir conexiones ocultas y predecir tendencias futuras tiene una larga historia. Conocido algunas veces como "descubrimiento de conocimientos en bases de datos", el término "minería de datos" no se acuñó sino hasta la década de 1990. Pero su base comprende tres disciplinas científicas entrelazadas: estadística (el estudio numérico de relaciones de datos), inteligencia artificial (inteligencia similar a la humana exhibida por software y/o máquinas) y machine learning (algoritmos que pueden aprender de datos para hacer predicciones). Lo que era antiguo es nuevo otra vez, ya que la minería de datos continúa evolucionando para igualar el ritmo del potencial sin límites del big data y poder de cómputo asequible. En la última década, los avances en el poder y la velocidad de procesamiento nos han permitido llegar más allá de las prácticas manuales, tediosas y que toman mucho tiempo al análisis de datos rápido, fácil y automatizado. Cuanto más complejos son los conjuntos de datos recopilados, mayor es el potencial que hay para descubrir insights relevantes. Los comerciantes detallistas, bancos, fabricantes, proveedores de telecomunicaciones y aseguradoras, entre otros, utilizan la minería de datos para descubrir relaciones entre todas las cosas, desde precios, promociones y demografía hasta la forma en que la economía, el riesgo, la competencia y los medios sociales afectan sus modelos de negocios, ingresos, operaciones y relaciones con clientes.

¿Entonces, por qué es importante la minería de datos? En primer lugar porque más información no significa necesariamente más conocimientos. Y en segundo lugar porque permite:

- Filtrar todo el ruido caótico y repetitivo en sus datos.
- Entender qué es relevante y luego hacer un buen uso de esa información para evaluar resultados probables.
- Acelerar el ritmo de la toma de decisiones informadas.

¿Quién la utiliza?

La minería de datos se sitúa en el corazón de esfuerzos analíticos en diversas industrias y disciplinas.

Comunicaciones:

En un mercado sobrecargado donde la competencia es cerrada, las respuestas se encuentran a menudo en los datos de sus consumidores. Las compañías de multimedia y telecomunicaciones pueden utilizar modelos analíticos para entender montañas de datos de clientes, ayudándoles así a predecir el comportamiento de sus clientes y ofrecer campañas altamente dirigidas y relevantes.

Seguros:

Con conocimientos analíticos, las compañías de seguros pueden resolver problemas complejos concernientes a fraude, cumplimiento, gestión de riesgo y separación de clientes. Las compañías han utilizado técnicas de minería de datos para asignar precios a productos con mayor eficacia en líneas de negocios y hallar nuevas formas de ofrecer productos competitivos a su base de clientes existente.

Educación:

Con vistas unificadas basadas en datos del progreso de los estudiantes, los educadores pueden predecir el desempeño de sus alumnos antes de que pongan un pie en el salón de clases – y desarrollar estrategias de intervención para mantenerlos en curso. La minería de datos ayuda a los educadores a acceder a datos de los estudiantes, a predecir niveles de logro y a detectar estudiantes o grupos de estudiantes que necesitan atención extra.

Manufactura:

La alineación de planes de suministro con pronósticos de demanda es esencial, como lo es también la detección temprana de problemas, garantía de calidad e inversión en equidad de marca. Los fabricantes pueden predecir el desgaste de activos de producción y anticipar su mantenimiento, lo cual puede maximizar el tiempo en operación y mantener la línea de producción acorde a lo programado.

Bancos:

Los algoritmos automatizados ayudan a los bancos a entender a su base de clientes y también los miles de millones de transacciones en el corazón del sistema financiero. La minería de datos ayuda a las compañías de servicios financieros a tener una mejor vista de los riesgos del mercado, a detectar el fraude en menos tiempo, a gestionar las obligaciones de cumplimiento de las regulaciones y a obtener retornos óptimos de sus inversiones en marketing.

Retail:

Grandes bases de datos de clientes contienen insights ocultos que le pueden ayudar a mejorar las relaciones con clientes, optimizar campañas de marketing y pronosticar ventas. A través de modelos de datos más precisos, las compañías detallistas pueden ofrecer campañas más enfocadas – y encontrar la oferta que tenga el mayor impacto en el cliente.

¿Cómo funciona?

La minería de datos, como una disciplina compuesta, representa diversos métodos o técnicas que se utilizan en diferentes capacidades analíticas que abordan una gama de necesidades organizacionales, hacen diferentes tipos de preguntas y utilizan diferentes niveles de aportación humana o reglas para llegar a una decisión.

Modelado descriptivo: Descubre similitudes o agrupaciones compartidas en datos históricos para determinar razones detrás del éxito o el fracaso, como la clasificación de clientes por preferencias de productos o sentimiento. Algunas técnicas de ejemplo incluyen:

- Agrupación en clústeres: agrupación de registros similares.
- Detección de anomalías: identificación de valores atípicos multidimensionales.
- Aprendizaje de reglas de asociación: detección de relaciones entre registros.
- Análisis de componentes principales: detección de relaciones entre variables.
- Agrupación por afinidad: agrupación de personas con intereses comunes o metas similares (por ejemplo, personas que compran X a menudo compran Y, y posiblemente Z).

Modelado predictivo: Este modelado llega más a fondo para clasificar eventos en el futuro o calcular resultados desconocidos – por ejemplo, el uso de evaluación de crédito para determinar la probabilidad de que una persona pague un préstamo. El modelado predictivo también ayuda a descubrir insights de cosas como la rotación de clientes, respuesta a campañas o coberturas por impago de créditos. Algunas técnicas de ejemplo incluyen:

Regresión: medida de la solidez de la relación entre una variable dependiente y una serie de variables independientes.

Redes neurales: programas de computadora que detectan patrones, hacen predicciones y aprenden.

Árboles de decisión: diagramas en forma de árbol en los que cada rama representa una incidencia probable.

Máquinas vectoriales de soporte: modelos de aprendizaje supervisado con algoritmos de aprendizaje asociados.

Modelado prescriptivo: Con el incremento de los datos no estructurados de la Web, campos de comentarios, libros, correo electrónico, PDFs, audio y otras fuentes de texto, la adopción de la minería de texto como disciplina relacionada con la minería de datos también ha crecido de manera considerable. Necesita la posibilidad de analizar, filtrar y transformar con éxito datos no estructurados para incluirlos en modelos predictivos para mejorar la precisión de las predicciones. Al final, no debe ver la minería de datos como una entidad independiente porque el preprocesamiento (preparación y exploración de datos) y el posprocesamiento (validación de modelos, calificación y monitoreo del desempeño de modelos) son igualmente esenciales. El modelado prescriptivo se centra en variables y restricciones internas y externas para recomendar uno o más cursos de acción – por ejemplo, determinar la mejor oferta de marketing para enviar a cada cliente. Algunas técnicas de ejemplo incluyen:

Analítica predictiva más reglas: creación de reglas que pasaría si/entonces a partir de patrones y predicción de resultados.

Optimización del marketing: simulación de la mezcla de medios más ventajosa en tiempo real para lograr el más alto ROI posible.

Fuente: https://www.sas.com/es_ar/insights/analytics/data-mining.html#dmimportance