

MINERIA DE DATOS

INTRODUCCION

La minería de datos es algo que empieza a presentarse en los años sesenta, para este entonces los estadísticos manejaban términos como “Data Fishing”, “Data Mining (DM)” o “Data Archaeology”, con la idea de encontrar correlaciones sin una definición previa de bases de datos. Para los principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro y otros, empezaron a consolidar los términos de Minería de Datos y KDD.

De este modo la minería de datos actualmente se conoce como el proceso de detectar la información procesable de los conjuntos grandes de datos. Utilizando el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no pueden detectarse mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos.

La minería de datos nos permite filtrar todo el ruido caótico y repetitivo en los datos, entender qué es lo relevante y para posteriormente hacer un buen uso de la información obtenida para evaluar los resultados probables, acelerando el ritmo de la toma de decisiones informadas.

Todo esto es posible mediante la aplicación de diversas técnicas como por ejemplo las clases, los grupos, las asociaciones y los patrones secuenciales, los cuales se explicaran más adelante.

DESARROLLO

El software de minería de datos permite analizar las relaciones y patrones en los datos de transacción almacenados sobre la base de consultas de los usuarios de composición abierta.

Existen varios tipos de software de análisis que están disponibles: estadísticos, de aprendizaje automático, y redes neuronales. En general, se buscan cualquiera de estos cuatro tipos de relaciones:

- 3 Clases: Los datos almacenados se utilizan para localizar datos en grupos predeterminados. Por ejemplo, una cadena de restaurantes podría minar datos de compra del cliente para determinar cuando los clientes los visitan y lo que normalmente ordenan. Esta información podría utilizarse para aumentar el tráfico al tener especiales del día.

- 3 Grupos: Los elementos de datos se agrupan de acuerdo a las relaciones lógicas o preferencias de los consumidores. Por ejemplo, los datos pueden ser extraídos para identificar segmentos de mercado o afinidades de los consumidores.
- 3 Asociaciones: Los datos pueden ser minados para identificar asociaciones. El ejemplo de cerveza-pañal es un ejemplo de minería asociativa.
- 3 Patrones secuenciales: Los datos minan para anticipar patrones de comportamiento y tendencias. Por ejemplo, un distribuidor especializado en sistemas al aire libre puede predecir la probabilidad de compra de una mochila basado en la compra de un consumidor de sacos de dormir y calzado de senderismo.

La tecnología de la información ha estado evolucionando a gran escala de forma independiente a comparación de las transacciones y sistemas de análisis; la minería de datos proporciona un enlace entre los dos. La minería de datos se compone de cinco elementos principales:

1. Extraer, transformar y cargar datos de transacciones en el sistema de almacenamiento de datos.
2. Almacenar y manejar los datos en un sistema de base de datos multidimensional.
3. Proporcionar acceso a datos para los analistas de negocios y profesionales de tecnologías de la información.
4. Analizar los datos por un software de aplicación.
5. Presentar los datos en un formato útil, como un gráfico o tabla.

Así como de diversos niveles de análisis como lo son:

Redes neuronales artificiales

Modelos predictivos no lineales que aprenden a través de la formación y se asemejan a redes neuronales biológicas en su estructura.

Algoritmos genéticos

Técnicas de optimización que usan procesos tales como combinación genética, mutación y selección natural en un diseño basado en los conceptos de evolución natural.

Árboles de decisión

Estructuras en forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen Árboles de Clasificación y Regresión (CART) y Detección Automática de Interacción (Chi Cuadrado CHAID).

Método del vecino más próximo

Una técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases de registro(s) k más similares a él en un conjunto de datos históricos (donde $k \geq 1$). A veces se llama técnica del vecino k -más cercano.

¿Cómo se genera?

La generación de un modelo de minería de datos forma parte de un proceso mayor que incluye desde la formulación de preguntas acerca de los datos y la creación de un modelo para responderlas, hasta la implementación del modelo en un entorno de trabajo. Este proceso se puede definir mediante los seis pasos básicos siguientes:

I. Definir el problema

Consiste en definir claramente el problema y considerar formas de usar los datos para proporcionar una respuesta para el mismo. Este paso incluye analizar los requisitos empresariales, definiendo el ámbito del problema, las métricas por las que se evaluará el modelo y los objetivos concretos del proyecto de minería de datos. También se deben 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.

II. Preparar los datos

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. 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, para identificar los orígenes de datos que son más precisos y determinar qué columnas son las más adecuadas para el análisis.

III. Explorar los datos

Se deben 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. Al explorar los datos para conocer el problema empresarial, puede decidir si el conjunto de datos contiene datos defectuosos para de esta manera poder inventar una estrategia para corregir los problemas u obtener una descripción más profunda de los comportamientos que son típicos de su negocio.

IV. Crear modelos

Se deben definir qué columnas de datos se usaran; para ello, se debe crear 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. 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. Es importante recordar que siempre que los datos cambian, debe actualizar la estructura y el modelo de minería de datos.

V. Explorar y validar modelos

Consiste en explorar los modelos de minería de datos que ha generado y comprobar su eficacia. 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. 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.

VI. Implementar y actualizar los modelos

Consiste en implementar los modelos que funcionan mejor en un entorno de producción. 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. Pudiendo realizar Use los modelos para crear predicciones que luego podrá usar para tomar decisiones comerciales, Crear consultas de contenido para recuperar estadísticas, reglas o fórmulas del modelo, Incrustar la funcionalidad de minería de datos directamente en una aplicación, etc.

Viendo hacia el futuro

Aunque la minería de datos es un término relativamente nuevo, la tecnología no lo es. Las compañías han utilizado equipos de gran alcance para tamizar a través de volúmenes de datos de escaneo de los supermercados y analizar los informes de investigación de mercado durante años.

Sin embargo, las continuas innovaciones en la capacidad de procesamiento, almacenamiento en disco, y software para estadística están aumentando drásticamente la precisión del análisis, mientras reducen el coste.

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. Cuantos 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.

La minería de datos se utiliza sobre todo hoy en día por las empresas con un enfoque fuerte en los consumidores – comunicaciones, comercio, finanzas, y las organizaciones de comercialización.

Esto permite que las empresas determinen las relaciones entre los factores "internos" como el precio, posicionamiento del producto, o las habilidades del personal, y factores "externos", como los indicadores económicos, la competencia, y la demografía de los clientes. Y, les permite determinar el impacto en las ventas, la satisfacción del cliente, y las ganancias corporativas. Por último, les permite "profundizar" en la información resumida para ver datos detallados de las transacciones.

Con la minería de datos, un minorista puede utilizar los registros de puntos de venta de compras de los clientes para enviar promociones específicas basadas en el historial de compras de un individuo. Minando datos demográficos de las tarjetas de comentarios o de garantía, el distribuidor podría desarrollar productos y promociones para atraer a segmentos específicos de clientes.

CONCLUSION

La minería de datos al permitirle a las grandes y medianas empresas ahorrar grandes cantidades de dinero, le brinda la apertura a nuevas oportunidades de negocios, contribuyendo a la buena toma de decisiones para que se generen mejores resultados

de la forma más rápida, eficaz y al menor costo. Todo esto es posible si general modelos descriptivos para poder explorar, comprender e identificar patrones que puedan causar un impacto en los resultados esperados.

BIBLIOGRAFIAS

<https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>

https://www.sas.com/es_mx/insights/analytics/data-mining.html

<https://www.tecnologias-informacion.com/mineria-de-datos.html>

<https://es.slideshare.net/janettejf/mineria-de-datos-3582262#:~:text=Es%20un%20mecanismo%20de%20explotaci%C3%B3n,de%20confianza%20para%20cada%20predicci%C3%B3n.>