

Analítica y tipos de aprendizaje automático

Mauricio Antonio Alonso
Moncada



Tipos de análisis de datos (Data Analysis)

Estudiaremos con mayor profundidad los siguientes tipos básicos de análisis de datos (Data Analysis):

- Análisis cuantitativo
- Análisis cualitativo
- Minería de datos (Data Mining)

Cada descripción incluye un ejemplo sencillo basado en el escenario de ventas de conos de helado usado en la descripción inicial del análisis de datos (Data Analysis).

Análisis cuantitativo

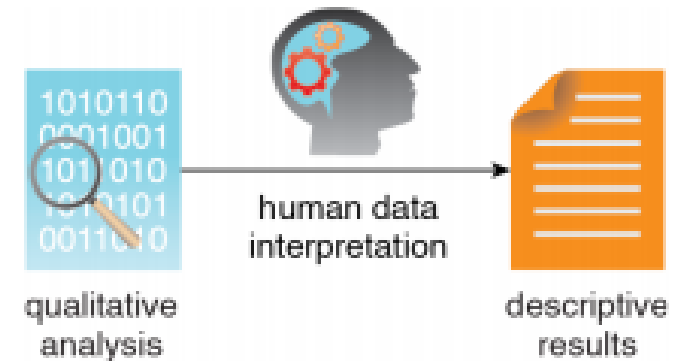
El análisis cuantitativo es una técnica de análisis de datos (Data Analysis) orientada a cuantificar patrones y correlaciones hallados en los datos. Esta técnica implica el análisis de un gran número de observaciones de un dataset con base en técnicas estadísticas. Debido al amplio tamaño de la muestra, los resultados pueden aplicarse de manera general a todo el dataset.

Los resultados del análisis cuantitativo son de naturaleza absoluta y, por lo tanto, pueden ser usados para realizar comparaciones numéricas. Por ejemplo, en un análisis cuantitativo de las ventas de helados, se puede encontrar que un aumento de 5 grados en la temperatura incrementa las ventas en un 15%.

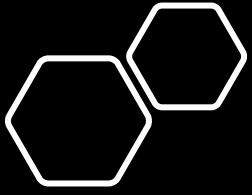


Análisis cualitativo

El análisis cualitativo es una técnica de análisis de datos (Data Analysis) orientada a describir cualidades de varios datos por medio de palabras. En contraste con el análisis de datos (Data Analysis) cuantitativo, esto implica analizar una pequeña muestra con mayor profundidad.



Los resultados de este análisis no se pueden aplicar de forma general a todo un dataset debido al pequeño tamaño de la muestra. Tampoco pueden ser medidos numéricamente o usados para comparaciones numéricas. Por ejemplo, un análisis de las ventas de conos de helado puede indicar que las cifras de las ventas en mayo no fueron tan altas comparadas con el mes de junio. Los resultados del análisis solo muestran que las cifras “no fueron tan altas comparadas con”, mas no indican ninguna diferencia numérica.



Minería de datos (Data Mining)

Es una forma especializada de análisis de datos (Data Analysis) dedicada a los datasets grandes. En relación con el análisis de Big Data, la minería de datos (Data Mining) por lo general se refiere a técnicas automáticas basadas en software que filtran los datasets masivos para identificar patrones y tendencias.



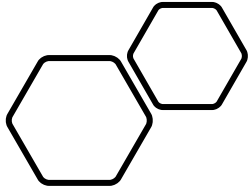
Específicamente, implica extraer patrones ocultos o desconocidos en los datos con la intención de identificar patrones antes desconocidos. La minería de datos (Data Mining) constituye la base para la analítica predictiva y la Inteligencia de negocios (BI).

Análisis y Analítica

El tiempo y el esfuerzo requeridos para llevar a cabo un análisis manual aumentan considerablemente con Big Data. Es posible que estas técnicas no brinden hallazgos exactos ni de manera oportuna debido al volumen, velocidad, y/o variedad de los datos. Estas ineficiencias potenciales se agravan aún más si se debe repetir el análisis.

Las herramientas de analítica pueden automatizar el análisis de datos (Data Analysis) usando tecnologías computacionales altamente escalables que aplican técnicas de análisis estadísticos cuantitativos automatizados, minería de datos (Data Mining) y aprendizaje automático (Machine Learning)

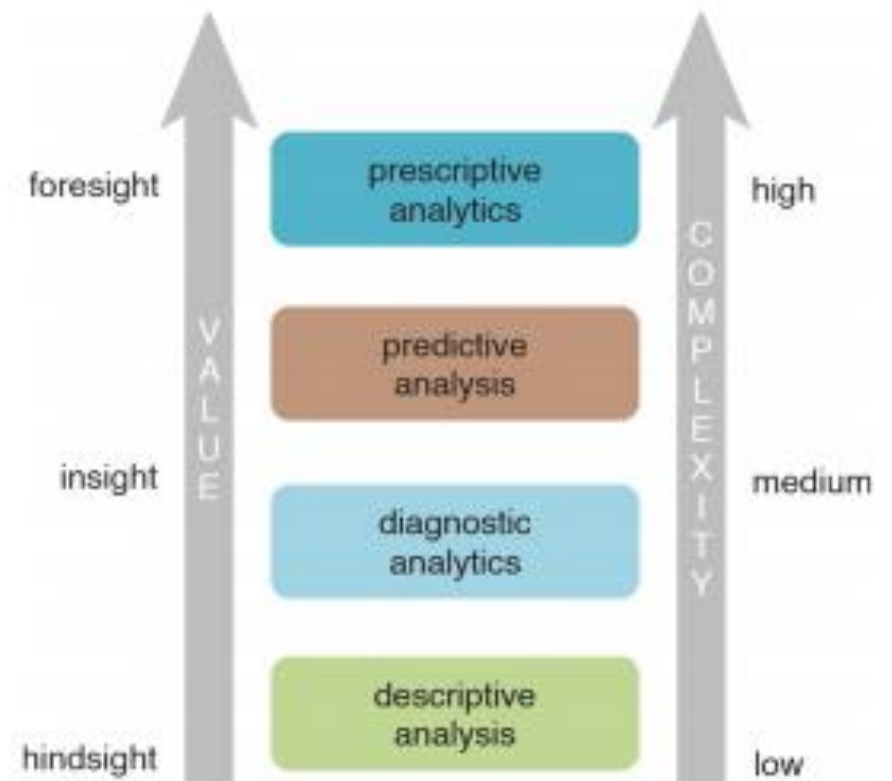




Tipos de Analítica

El término “analítica” se describió inicialmente en la sección Terminología y conceptos fundamentales. Esta sección explora más a fondo la analítica, describiendo a continuación cuatro tipos comunes de analítica:

- Analítica descriptiva
- Analítica diagnóstica
- Analítica predictiva
- Analítica prescriptiva





Analítica descriptiva

La analítica descriptiva se ejecuta para responder preguntas sobre eventos que ocurrieron. Algunas preguntas de ejemplo pueden ser:

- ¿Cuáles son los datos de las ventas de los últimos 12 meses?
- ¿Cuántas llamadas de soporte técnico fueron recibidas y categorizadas según la gravedad y ubicación geográfica?
- ¿Cuál es la comisión mensual que gana cada agente de ventas?

Analítica descriptiva (Cont)

Alrededor del 80% de la analítica es de naturaleza descriptiva. En términos de valor, la analítica descriptiva proporciona un valor mínimo y requiere un conjunto relativamente básico de habilidades. A menudo, la analítica descriptiva se ejecuta mediante reportes o tableros de control (Dashboards) especializados.

Por lo general, los reportes son de naturaleza estática y muestran datos históricos que son presentados en forma de grillas de datos o gráficos. Las consultas son ejecutadas en los sistemas de OLTP o en los datos obtenidos a partir de una variedad de otros sistemas de información, tales como CRM Y ERP.



Analítica diagnóstica

La analítica diagnóstica tiene como objetivo determinar la causa de un fenómeno que ocurrió en el pasado, usando preguntas que se enfocan en la razón del evento.

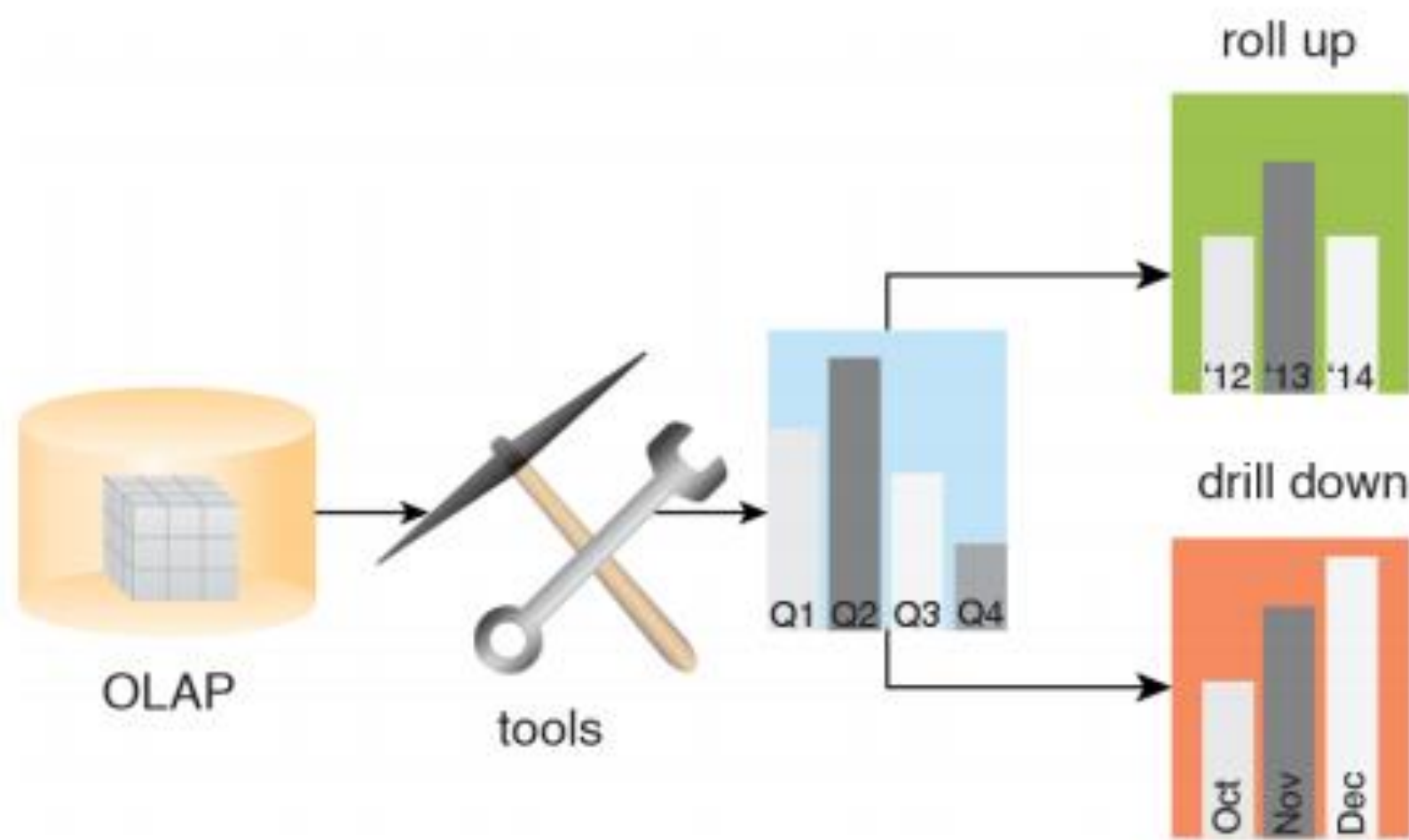
Algunas preguntas de ejemplo pueden ser:

- ¿Por qué las ventas del segundo trimestre fueron menores que las de primer trimestre?
- ¿Por qué se han recibido más llamadas de soporte técnico de la región este que de la región oeste?
- ¿Por qué hubo un incremento en las tasas de readmisión de pacientes en los últimos tres meses?



Analítica diagnóstica (cont)

Se considera que la analítica diagnóstica proporciona más valor que la analítica descriptiva, y que requiere un conjunto de habilidades más avanzadas. Normalmente, también requiere recopilar datos a partir de múltiples fuentes y almacenarlos en una estructura que se presta para realizar drill-downs y roll-ups



Analítica predictiva



La analítica predictiva se ejecuta en un intento por determinar el resultado de un evento que podría ocurrir en el futuro. Las preguntas normalmente se formulan usando una lógica condicional qué tal sí, como en los siguientes ejemplos:

- ¿Cuáles son las probabilidades de que un cliente incurra en el incumplimiento de un préstamo si él no ha hecho el pago mensual?
- Si se administra el medicamento B en vez del medicamento A, ¿cuál será la tasa de supervivencia del paciente?
- Si un cliente compra los Productos A y B, ¿qué posibilidades hay de que también compre el Producto C?

Analítica predictiva (cont)

La analítica predictiva intenta predecir el resultado de un evento. Las predicciones se hacen con base en patrones, tendencias y excepciones encontradas en datos históricos y actuales. Este análisis permite identificar riesgos y oportunidades.

La analítica predictiva implica el uso de datasets grandes compuestos por datos internos y externos, además de técnicas estadísticas, de análisis cuantitativo, de aprendizaje automático (Machine Learning) y de minería de datos (Data Mining). Se considera que este tipo de analítica proporciona más valor y requiere un conjunto de habilidades más avanzadas que las analíticas descriptiva y diagnóstica. Por lo general, las herramientas usadas abstraen particularidades estadísticas, proporcionando una interfaz de front-end fácil de usar



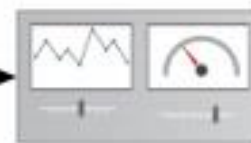
OLAP



textual
data



analytics



predictive
analytics

Analítica prescriptiva

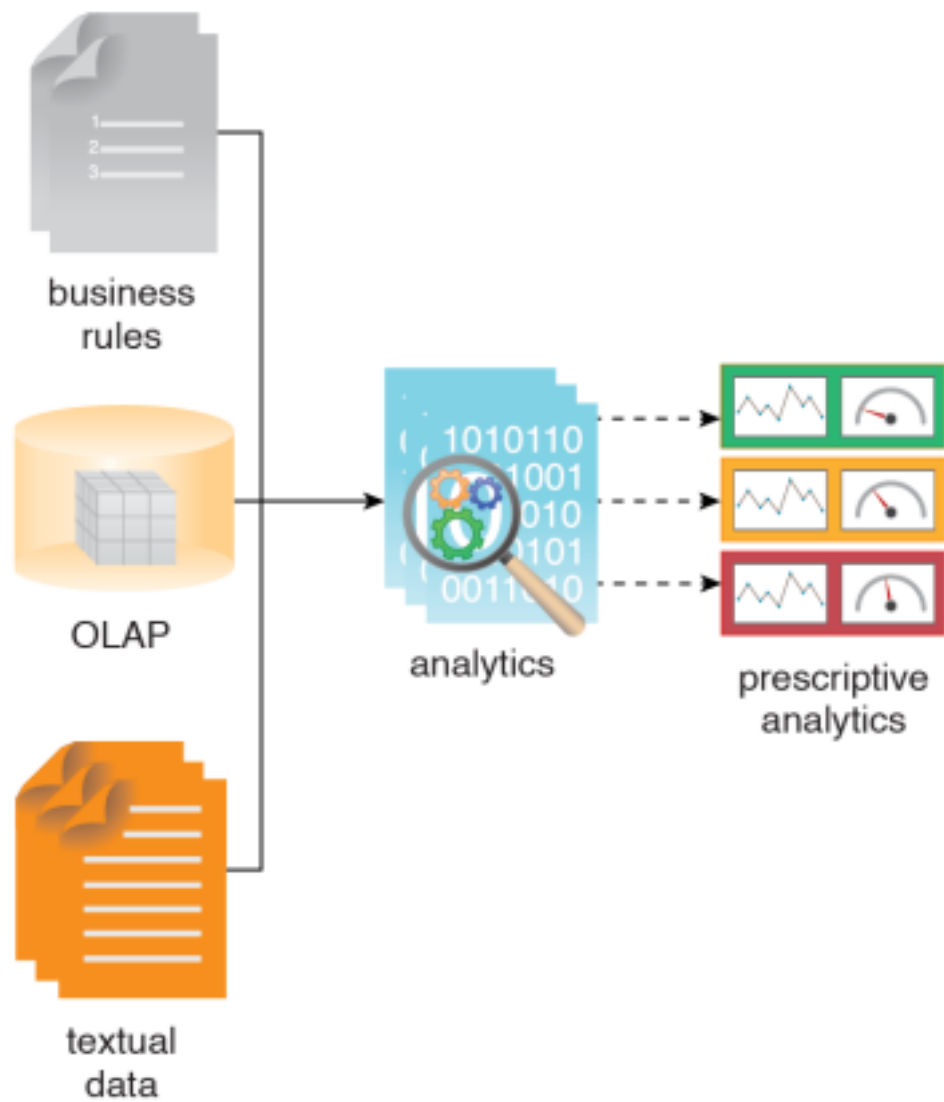
La analítica prescriptiva está basada en los resultados de la analítica predictiva, al indicar acciones que se deberían realizar. Esta analítica se enfoca en qué opción indicada se debe seguir y en por qué y cuándo se debería seguir, con el fin de obtener una ventaja o mitigar un riesgo. Algunas preguntas de ejemplo pueden ser:

- Entre tres opciones de medicamentos, ¿cuál ofrece los mejores resultados?
- ¿Cuándo es el mejor momento para comercializar una mercancía particular?

Analítica prescriptiva (cont)

La analítica prescriptiva proporciona más valor que cualquier otro tipo de analítica, y en consecuencia, requiere un conjunto de habilidades más avanzadas, además de software y herramientas especializadas. Se calculan varios resultados y se sugiere el mejor plan de acción para cada uno de ellos. Se hace un cambio del enfoque explicativo al enfoque consultivo, y puede incluir la simulación de varias situaciones.

La analítica prescriptiva incorpora datos internos, incluidos datos de ventas actuales e históricos, información de los clientes, datos del producto, normas comerciales y también datos externos, incluidos datos de social media, datos climatológicos y datos demográficos. La analítica prescriptiva implica el uso de normas comerciales y grandes cantidades de datos internos y/o externos para simular los resultados y prescribir el mejor plan de acción



Analítica prescriptiva (cont)

Aprendizaje automático (Machine Learning)

El aprendizaje automático (Machine Learning) es el proceso de enseñar a las computadoras a aprender a partir de datos existentes y a aplicar el conocimiento adquirido para formular predicciones sobre datos desconocidos. Esto implica identificar patrones en los datos de entrenamiento y clasificar datos nuevos y no mostrados con base en patrones conocidos. Los algoritmos de aprendizaje automático (Machine Learning) también permiten ajustar los comportamientos utilizando un loop de retroalimentación, a la vez que funcionan con datasets nuevos. Normalmente, estos algoritmos se dividen en los siguientes dos tipos:

- Aprendizaje supervisado
- Aprendizaje no supervisado

Tipos de aprendizaje automático (Machine Learning)

Primero, los datos de muestra son enviados al algoritmo de aprendizaje supervisado, donde ya se conocen las categorías de los datos. Con base en los datos introducidos, el algoritmo comprende qué datos corresponden a qué categoría. Posteriormente, el algoritmo puede aplicar el comportamiento que aprendió para categorizar los datos desconocidos.

Tipos de aprendizaje automático (Machine Learning)

En un algoritmo de aprendizaje no supervisado, no se conocen las categorías de los datos y no se envía ningún dato de muestra. En vez de eso, el algoritmo intenta categorizar los datos agrupándolos según atributos similares.

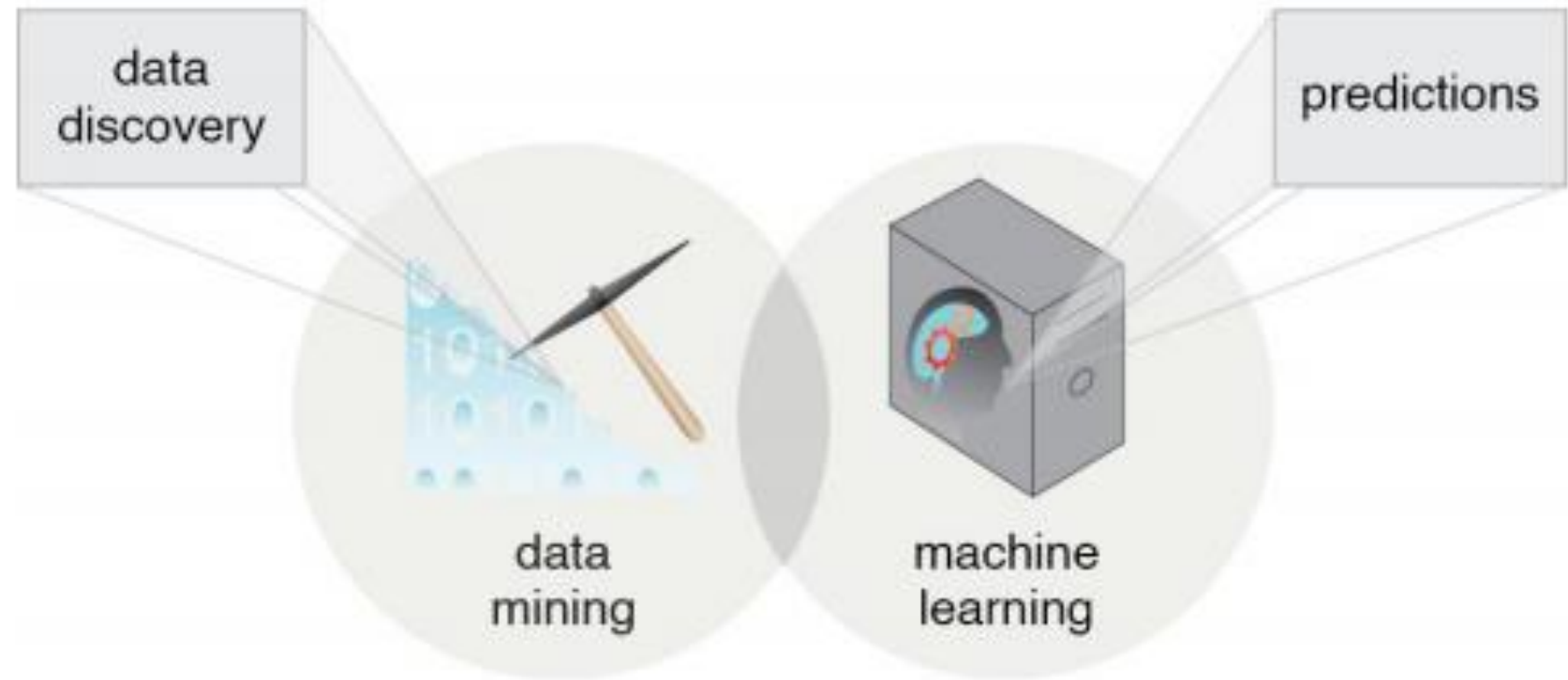
Comparación entre el aprendizaje automático (Machine Learning) y la minería de datos (Data Mining)

A pesar de que la minería de datos (Data Mining) y el aprendizaje automático (Machine Learning) están estrechamente relacionados, tienen diferencias notables. Considerando que la minería de datos (Data Mining) encuentra patrones ocultos y relaciones basadas en atributos de datos antes desconocidos, el aprendizaje automático (Machine Learning) hace predicciones categorizando datos basados en patrones conocidos.



Comparación entre el aprendizaje automático (Machine Learning) y la minería de datos (Data Mining)

La minería de datos (Data Mining) puede emplear algoritmos de aprendizaje automático (Machine Learning) —como el aprendizaje no supervisado— para extraer atributos antes desconocidos. Esto se logra mediante la categorización de datos, lo cual lleva a la identificación de patrones. El aprendizaje automático (Machine Learning) puede usar el resultado de la minería de datos (Data Mining) o los patrones identificados para clasificarlos aún más mediante el aprendizaje supervisado



Preguntas





TALLER