

MINERA DE DATOS

El proyecto de minería de datos comienza con un plan bien definido de inteligencia comercial. Los analistas de la empresa definen el problema a resolver y el objetivo concreto de empresa que se desea cumplir. y las funciones de minería cuanto mejor sea esta formulación inicial más claras serán las directrices acerca de los datos y las funciones de minería que se utilizan para conseguir los resultados deseados.

MODELO PREDICTIVO

es un término paraguas para referirnos al conjunto de procesos que implican aplicar diferentes técnicas computacionales con el objetivo de realizar predicciones sobre el futuro basándonos en datos pasados. La variedad de técnicas empleadas incluye minería de datos, modelado reconocimiento de patrones, graph analytics, etc.

TIPOS

CLASIFICACIÓN

Permiten predecir la pertenencia a una clase. Por ejemplo si tratamos de clasificar entre nuestros clientes quiénes son más propensos a abandonar. Para ello se establecen variables de entrada como el riesgo de crédito, respuestas a comunicaciones. Los resultados del modelo son binarios, o un sí o un no (en forma de 0 y 1) con su grado de probabilidad.

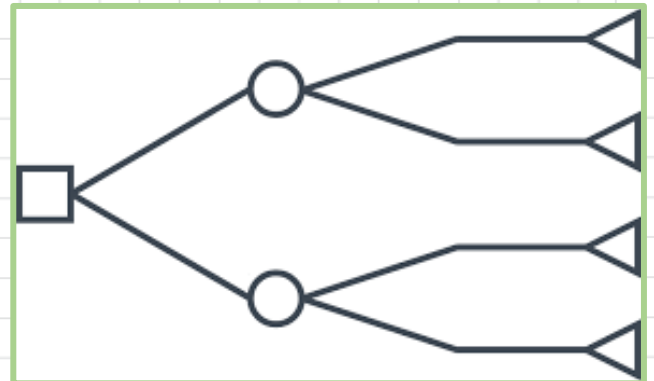
REGRESIÓN

Nos permiten predecir un valor. Por ejemplo, cuál es el beneficio de un determinado cliente (o segmento) en los próximos meses.

TÉCNICAS

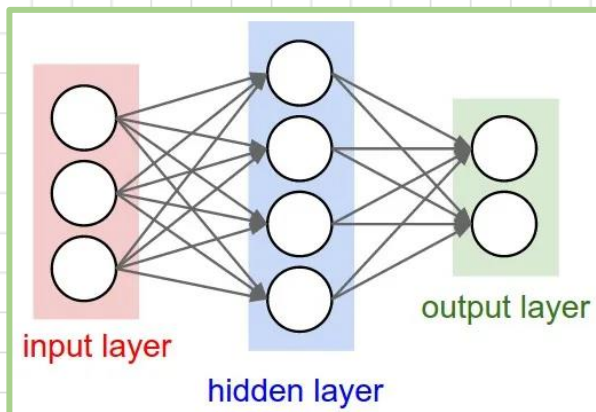
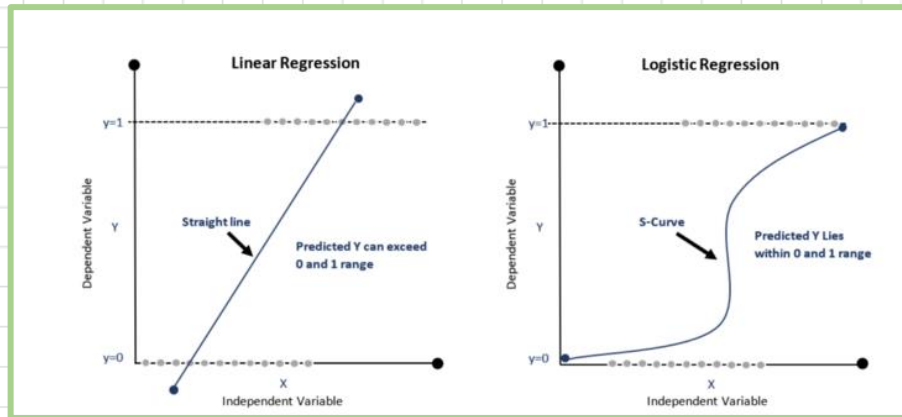
ÁRBOLES DE DECISIÓN

Se utilizan para desarrollar sistemas de clasificación que predicen o clasifican observaciones futuras basándose en un conjunto de reglas de decisión. Si dispone de datos divididos en clases que le interesan (por ejemplo préstamos de alto riesgo frente a préstamos de bajo riesgo, suscriptores frente a no suscriptores, votantes frente a no votantes o tipos de bacterias), puede usar los datos para generar reglas que pueda usar para clasificar casos antiguos o recientes con la máxima precisión.



REGRESIÓN LINEAL Y LOGÍSTICA

La Regresión Lineal, se utiliza para proyectos de es por ejemplo, queremos predecir el precio de una casa o el decir cuándo queremos predecir un valor sueldo de una persona de acuerdo a los años de experiencia. Por su parte el algoritmo de Regresión Logística se utiliza para proyectos de clasificación, si queremos conocer si una persona vio o no el naufragio del Titanic, o si una acción de la bolsa de valores va a subir o no.



REDES NEURONALES

Consisten en técnicas sofisticadas capaces de modelar relaciones extremadamente complejas. Se han hecho populares porque son muy potentes pero flexibles al mismo tiempo. Tienen la capacidad de manejar relaciones no lineales de los datos, lo que lo hace muy interesante cuantos más datos manejemos, ideal para el análisis de big data. Otras veces se usan simplemente para confirmar descubrimientos de otras técnicas más sencillas como las regresiones o los árboles de decisión.

MODELO DESCRIPTIVO

La analítica descriptiva examina los datos y analiza los acontecimientos pasados para saber cómo abordar el futuro. La analítica descriptiva examina el rendimiento pasado y entiende ese rendimiento al extraer datos históricos para buscar las razones detrás del éxito o el fracaso del pasado. Casi todos los informes de gestión, tales como ventas, marketing, operaciones y finanzas utilizan este tipo de análisis post-mortem.

TIPOS

GRÁFICA

Su fin es representar de forma gráfica la información contenida en los datos.
Existen tres tipos básicos de gráficos: Sectores, Barras e Histograma.
Dependiendo del tipo de variable se utilizará un tipo de gráfico pero no otro.

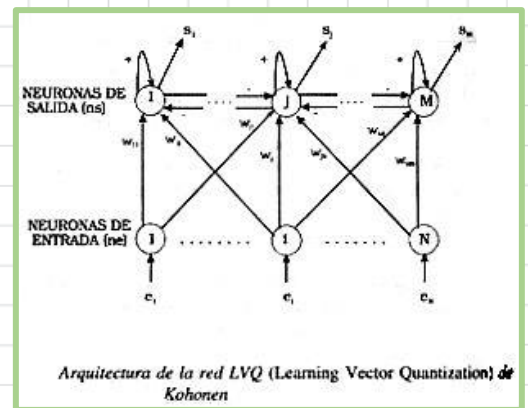
ANALITICAS

Su fin es resumir en forma de "parámetros" o valores de ciertas características de las variables toda la información contenida en los datos.

TECNICAS

EL MODELO DE KOHONEN

Se trataba de un modelo de red neuronal con capacidad para formar mapas de características de manera similar como ocurre en el cerebro. En éste hay neuronas que se organizan en muchas zonas, de forma que las informaciones captadas del entorno a través de los órganos sensoriales se representan internamente en forma de mapas bidimensionales.



K) MEDIDAS

Nos indican en torno a qué valor (centro) se distribuyen los datos. Las más comunes son las siguientes:

MODA: Es el valor que tiene mayor frecuencia absoluta. Variables cualitativas y cuantitativas.

MEDIANA: Es el valor que ocupa el lugar central de todos los datos cuando estos están ordenados de menor a mayor. Variables cuantitativas.

MEDIANA ARITMÉTICA: Es el valor obtenido al sumar todos los datos y dividir el resultado entre el número total de datos.

MEDIDAS DE POSICIÓN: Nos dan un valor promedio o característica representativa de la muestra.

MEDIDAS DE DISPOSICIÓN: Sirven para medir el grado de representativo de las medidas de posición central respecto a los datos.

Unidades básicas del SI (Sistema Internacional de unidades)		
Nombre	Símbolo	Magnitud básica
amperio	A	intensidad de corriente eléctrica
candela	cd	intensidad luminosa
kelvin	K	temperatura termodinámica
kilogramo	kg	masa
metro	m	longitud
mol	mol	cantidad de sustancia
segundo	s	tiempo, duración

VECINOS MAS PROXIMOS

Es un método sencillo de clasificación y regresión es el de los k vecinos más cercanos

Es un método no paramétrico. Se usará la notación $NN(k, x_{\phi})$ para denotar el conjunto de los k vecinos más cercanos al objeto x_{ϕ} .

