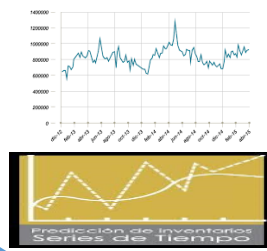
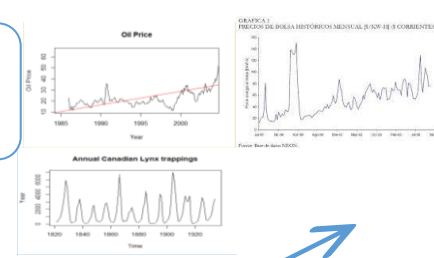


Las reglas de asociación nos permiten contar las combinaciones de artículos o ítems que ocurren con mayor frecuencia en una base de datos.
Base de datos transaccional:
Lista
Representación vertical
Representación horizontal



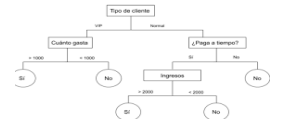
Es un conjunto de valores observados durante un periodo determinado de tiempo. este periodo está dividido en intervalos que se encuentran separados de manera regular.



Es muy utilizado en la industria de ventas al por menor, y también en el dominio de la medicina.
 Características: el orden importa.
 Objetivo: Encontrar patrones secuenciales.



Es un diagrama que representa en forma secuencial condiciones y acciones.
 • Son fáciles de construir, interpretar y visualizar.



Técnicas de minería de datos

Series de tiempo

Patrones secuenciales

Árboles de Decisión

K-medias

Visualización

Regresión Lineal

Métricas de evaluación

Máquinas de soporte vectorial

Detección de outliers

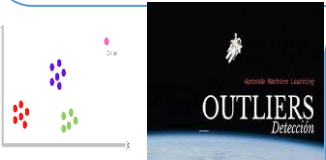
Regresión Bayesiana

Reglas de asociación

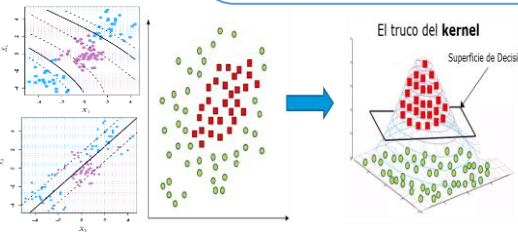
Proceso de inferencia
 Las hipótesis científicas se expresan a partir de distribuciones de probabilidad formuladas a partir de la observación de los datos.
 Ventajas:
 Permite ajustar modelos complejos no abordables por método de frecuencias.
 Permite realizar propagaciones de error.
 Minimiza el uso de límites arbitrarios para tomar decisiones.



Valores que se "escapan al rango en donde se concentran la mayoría de muestras".
Significan varias cosas:
 • Error
 • Límites
 • Punto de interés

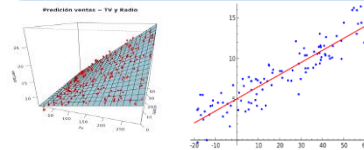


Son un conjunto de algoritmos de aprendizaje supervisados que desarrollan métodos relacionados con los problemas de clasificación y regresión

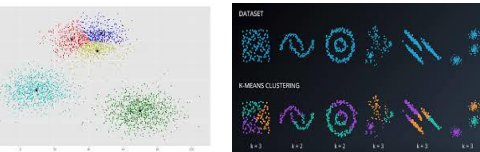


Estimar la precisión de la generalización de un modelo sobre los datos futuros .
Clasificación:
 • Máquinas de soporte vectorial
 • Árboles de decisión
Se divide en:
 • Precisión
 • Regresión
 • Asociación
 • Agrupamiento

Es un campo de estudio que enfatiza la relación estadística entre 2 variables continuas conocidas como variable de predicción y respuesta.
Regresión lineal simple
 Este modelo solo esta conformado por dos variables estadísticas llamadas x y y



Es el proceso de representación de datos, en formato gráfico, de una manera clara y eficaz
Ventajas:
 • Ayuda a tomar decisiones acertadas.
 • Ahorro de tiempo.
 • Optimiza la colaboración
 • funciones de autoservicio para los usuarios.
 • reducción de la carga de trabajo del equipo de tecnología



Es un algoritmo no supervisado de clustering, se utiliza cuando tenemos demasiados datos sin etiqueta para con esto encontrar cada grupos entre los datos crudos
 Sirve para tener escalabilidad con la cantidad de datos. el algoritmo es de los más usados para encontrar grupos ocultos sobre un conjunto de datos no etiquetados

