

Una BDD transaccional puede ser representada como una lista, de manera vertical u horizontal.

La manera mas eficiente es la representación vertical, ya que solo ocupa dos columnas.

A priori es uno de los algoritmos más utilizados para obtener reglas de asociación.

Tienen como objetivo encontrar relaciones dentro de un conjunto de transacciones

Reglas de asociación

Metricas de interes:  
- Soporte (frecuencia relativa)  
- Confianza (probabilidad)  
- Lift (refleja la probabilidad)

Tipos de outliers: casos atipicos, extraordinarios y observaciones únicas

Técnicas de minería de datos

El RSE son los datos que el modelo no puede explicar debido a falta de información y sirve para reducir el error cuando se agregan más variables.

Se puede usar esta técnica siempre que los datos sean veridicos

Outliers

Regresión lineal

Existen los outliers univariados y multivariados

Metodos de detección: desviación estandar, boxplots y DBScan

Puede ocurrir debido a errores o circunstancias extraordinarias

Se busca una variable aleatoria y la manera en que se ve influenciada

Se da apartir de la siguiente formula:  
$$Y = \alpha + \beta * X + E$$

A partir de los datos historicos se puede realizar una predicción

Se busca una variable aleatoria y la manera en que se ve influenciada

# Técnicas de minería de datos

## Clustering

## Análisis predictivo

Algunas de sus aplicaciones seria en el marketing para ayudar a segmentar clientes

Tiene como objetivo identificar la similitud (alta o baja) entre clústers

Extrae información existente en los datos y la utiliza para predecir tendencias y patrones

Puede utilizarse para predecir el pasado, presente o futuro

Es una técnica que ideantifica agrupaciones de manera automatica.

Debe ser escalable, manejar varios aributos, independiente del orden de entrada de datos e interpretable

Consiste en la tecnologia que aprende de la experiencia para predecir el futuro

Para llevarse acabo requiere una gran cantidad de datos

Dependiendo del resultado a obtener se puede utilizar la distancia euclídea o manhattan

Existen dos tipos de clustering: partición y jararquico. Aunque se utiliza mas el segundo por practicidad

Gracias al avance tencologico, se pueden hacer predicciones incluso a partir de los datos que genera una persona

Regresión lineal, árboles de clasificación y regresión, redes neuronales, naive bayes, entre otras, son las técnicas mas utilizadas para analisis predictivo

Aún no tenemos claro  
cual técnica  
utilizaremos, puesto  
que no hemos elegido  
una base de datos, pero  
la de regresión y  
clustering me llaman la  
atención por la variedad  
de técnicas que poseen