



Aprendizaje no supervisado

(Fundamentos y aplicaciones)

M.Sc. Angelo Jonathan Diaz Soto

2025

Data Science – Business Intelligence – Big Data – Machine Learning – Artificial Intelligence – Innovation and
(+51) 976 760 803
Technology www.datayanalytics.com info@datayanalytics.com

- ¿Qué es el Aprendizaje no Supervisado? ■

- Tipos de Aprendizaje no Supervisado ■

- Algoritmos de Aprendizaje no Supervisado ■

- Tipos de

- algoritmos de agrupamiento ■ Algoritmo de agrupamiento K-means ■ Market Basket Analysis

- Reglas de asociación

- ¿Que es RFM ?



¿Qué es el Aprendizaje no Supervisado?



- El aprendizaje no supervisado es una de las formas en que el Machine Learning (ML) "aprende" los datos.
- El aprendizaje no supervisado no tiene datos sin etiquetar que el algoritmo tiene que intentar entender por sí mismo. ■ El



una de
Machine

datos.

supervisado
etiquetar que
que intentar
mismo. ■ El

aprendizaje supervisado es en el que se etiquetan los conjuntos de datos para que haya una clave de respuestas con la que la máquina pueda medir su precisión.

■ Si Machine Learning fuera un niño que aprendiera a andar en bicicleta, el aprendizaje supervisado es el padre que corre detrás de la bicicleta y la sostiene en posición vertical. El aprendizaje no supervisado consiste en entregar la bicicleta, darle palmaditas en la cabeza al niño y decirle “buena suerte”.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

¿Qué es el Aprendizaje no Supervisado?



■ Los
Aprendizaje
inferen

algoritmos de
no
Supervisados
patrones de un

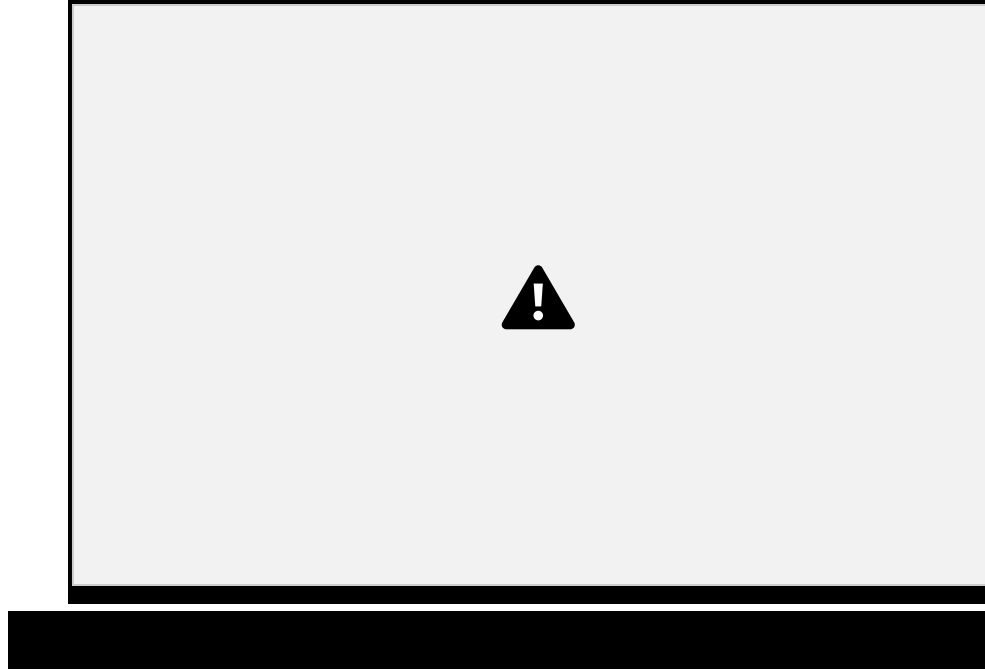
conjunto de datos sin referencia a resultados conocidos o etiquetados.

- ❑ Los métodos de Aprendizaje no Supervisado no se pueden aplicar directamente a un problema de regresión o clasificación porque no tiene idea de cuáles pueden ser los valores de los datos de salida, lo que hace imposible que entrene el algoritmo.
- ❑ El aprendizaje sin supervisión puede utilizarse para descubrir la estructura subyacente de los datos.
- ❑ Los algoritmos de Aprendizaje no Supervisados se utilizan para agrupar los datos no estructurados según sus similitudes y patrones distintos en el conjunto de datos.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

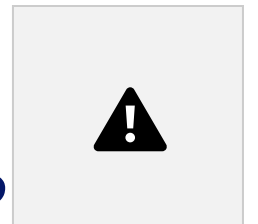
¿Qué es el Aprendizaje no Supervisado?





(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

¿Por qué es importante el Aprendizaje no Supervisado?



■ El
todo tipo de



etiquetados.

Aprendizaje no Supervisado encuentra
patrones
desconocidos en los datos.

- Los métodos no supervisados te ayudan a encontrar características que pueden ser útiles para la categorización.
- Es más fácil obtener datos no etiquetados que los datos

¿Cómo funciona el Aprendizaje no Supervisado?



- Los algoritmos de Aprendizaje no Supervisado manejan datos sin entrenamiento previo, es una función que hace su trabajo con los datos a su disposición.
- Los algoritmos no supervisados funcionan con datos no etiquetados. Su propósito es la exploración.

Los algoritmos de Aprendizaje no Supervisado están acostumbrados:

- 1 Explorar la estructura de la información y detectar patrones distintos. Extraer ideas valiosas. 2

Tipos de Aprendizaje no Supervisado

1. Agrupamiento:

El agrupamiento es un concepto importante cuando se trata de Aprendizaje no Supervisado. Se trata principalmente de encontrar una estructura o patrón en una colección de datos no categorizados. Los algoritmos de agrupamiento o clústeres, como se le conoce en inglés, procesarán los datos y encontrarán grupos o clústeres naturales si existen en los datos.





(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Tipos de Aprendizaje no Supervisado



2. Asociación:



Las reglas de asociación te permiten establecer asociaciones entre objetos de datos dentro de grandes bases de datos. Esta técnica no supervisada trata de descubrir relaciones interesantes entre variables en grandes bases de datos. Por ejemplo, las personas que compran una casa nueva tienen más probabilidades de comprar muebles nuevos.



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Tipos de Aprendizaje no Supervisado

3. Reducción de dimensionalidad:

La reducción de dimensionalidad como método de aprendizaje no supervisado tiene como objetivo reducir el número de variables a

tener en cuenta en el análisis.



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Algoritmos de Aprendizaje no Supervisado

■ Métodos de

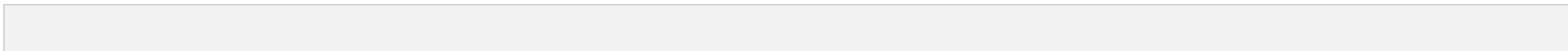


agrupamiento(K-mean, DBSCAN,etc.) ☐

Análisis de componentes principales

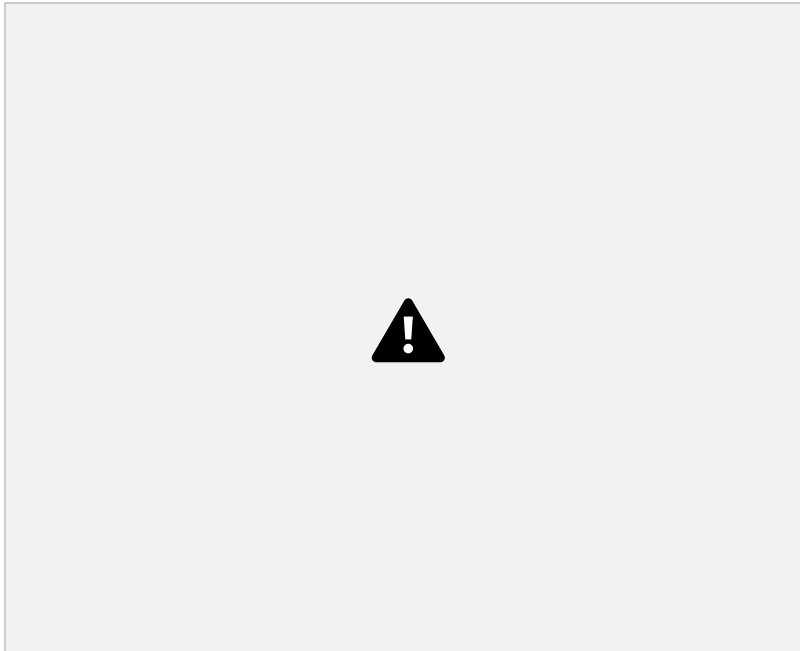
☐ Reglas de asociación

☐ Redes neuronales



Algoritmos de agrupamiento

El clustering o agrupación es una técnica de



inteligencia artificial dentro del abanico de Aprendizaje No Supervisado. Esto es así debido a que de antemano no se conocen las clases o el número de clases que se tendrán como resultado de la agrupación de dichos datos.





(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Algoritmos de agrupamiento

■ El análisis

clúster es una técnica multivariante cuya idea básica es clasificar objetos formando grupos/conglomerados (clusters) que sean lo más homogéneos posible dentro de si mismos y heterogéneos entre sí.

■ Este agrupamiento se basa en la idea de distancia o similitud entre las observaciones y la obtención de

dichos clusters depende del criterio o distancia considerados.

- El análisis clúster es una tarea de clasificación.





Aplicación del análisis cluster



- Segmentación de Mercados: Agrupamiento de consumidores de acuerdo a sus preferencias de atributos

Comprender el comportamiento de los compradores: Los consumidores con similares comportamientos/características son agrupados juntos.

- Identificar oportunidades de nuevos productos: Los clusters de similares marcas/productos pueden ayudar a identificar competidores/oportunidades de mercado



Objetivo en un análisis cluster

Unos buenos clusters deben cumplir dos objetivos:

- ☐ Minimizar distancia dentro de cada cluster.
- ☐ Maximizar la distancia

entre los diferentes



clusters.



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Tipos de algoritmos de agrupamiento



Existen

diferentes tipos de algoritmos de agrupamiento que manejan todo tipo de datos únicos.



1. Basado en densidad: En el agrupamiento basado en densidad, los datos se agrupan por áreas de altas concentraciones de puntos de datos rodeadas por áreas de bajas concentraciones de puntos de datos. Básicamente, el algoritmo encuentra los lugares que son densos en puntos de datos y los llama grupos.

2. Basado en la distribución: Con un enfoque de agrupamiento basado en la distribución, se considera que todos los puntos de datos forman parte de un grupo según la probabilidad de que un punto pertenezca a un grupo determinado.

Tipos de algoritmos de agrupamiento



3. Basado en Centroides: El agrupamiento basado en centroides es la que probablemente has escuchado más. Es algo sensible a los parámetros iniciales que le das, pero es rápida y eficiente.

4. Basado en Jerarquías: El agrupamiento basado en jerarquías se utiliza normalmente en datos

jerárquicos, como los que obtendrías de la base de datos de una empresa o de taxonomías. Construye un árbol de grupos para que todo esté organizado de arriba hacia abajo.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Tipos de algoritmos de agrupamiento



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Los Ocho Mejores Algoritmos de Agrupamiento

Algoritmo de agrupamiento K-means





Algoritmo de agrupamiento DBSCAN

2

Algoritmo de Mezcla Gaussiana

3

Algoritmo BIRCH

4

Algoritmo de agrupamiento por Propagación de Afinidad

5

Algoritmo de agrupamiento de Desplazamiento Medio

6

Algoritmo OPTICS

7



de agrupamiento K-means



- K-means es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características.

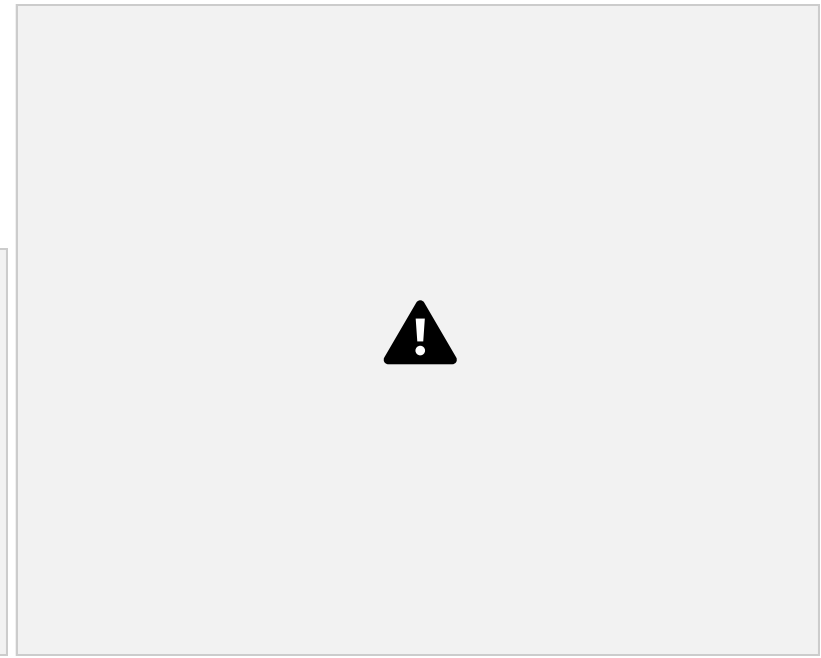
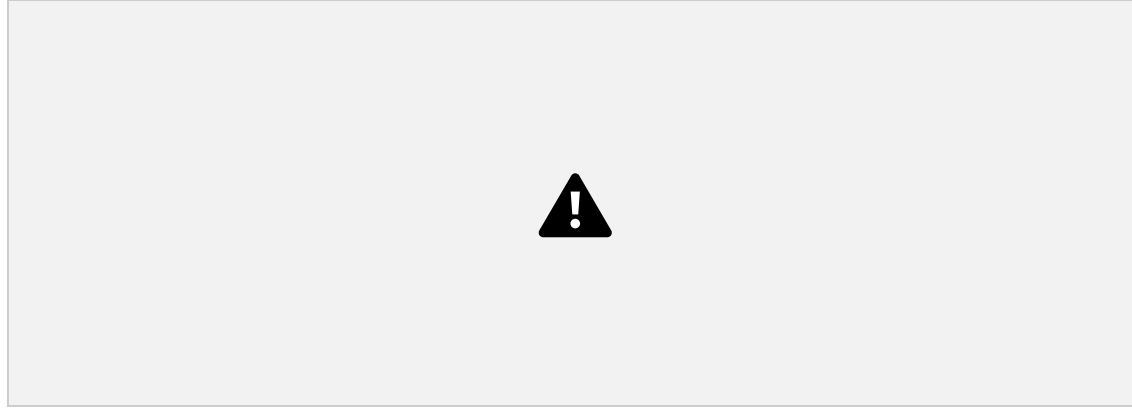
- El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Algoritmo de agrupamiento K-means

El algoritmo k-means resuelve un problema de optimización, siendo la función a optimizar (minimizar) la suma de las distancias





cuadráticas de cada objeto al centroide de su cluster.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Algoritmo de agrupamiento K-means

El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster. Se suele usar la distancia cuadrática.



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Algoritmo de agrupamiento K-means

El algoritmo consta de tres pasos:

1. Inicialización: una vez escogido el número de grupos, k , se establecen

k centroides en el espacio de los datos, por ejemplo, escogiendo aleatoriamente.

2. Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroide más cercano.
3. Actualización centroides: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

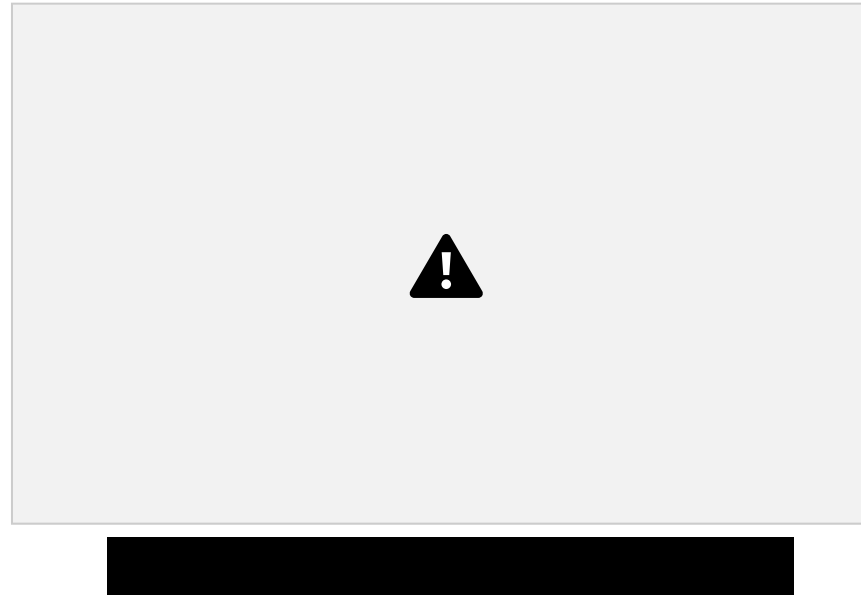
Elección del valor de K

- Una forma común de encontrar el valor óptimo de k es utilizar el método del codo. Para hacerlo, se grafican los valores de k junto con la suma de los errores cuadrados (SSE) para cada valor de k. SSE es la suma de la distancia de cada punto al

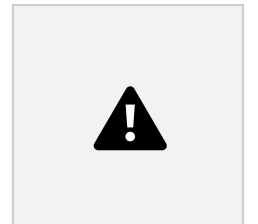


centroide de su cluster. A medida que aumenta k , el SSE disminuye, pero a un ritmo cada vez menor.

- El punto en el que el SSE disminuye más lentamente se conoce como el codo y es el punto óptimo para elegir k .



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Elección del valor de K

- Otra forma de encontrar el valor óptimo de k es utilizar técnicas de validación cruzada, como la validación cruzada de k -fold.
- Esto implica dividir el conjunto de datos en k subconjuntos y entrenar el algoritmo k -means k veces, cada vez utilizando un subconjunto diferente.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Casos de Uso de K-Means

El algoritmo de Clustering K-means es uno de los más usados para encontrar grupos ocultos, o

sospechados en teoría sobre un conjunto de datos no etiquetado.

Algunos casos de uso son:

- ❑ Segmentación por Comportamiento: relacionar el carrito de compras de un usuario, sus tiempos de acción e información del perfil.
- ❑ Categorización de Inventario: agrupar productos por actividad en sus ventas.
- ❑ Detectar anomalías o actividades sospechosas: según el comportamiento en una web reconocer un troll -o un bot- de un usuario normal.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

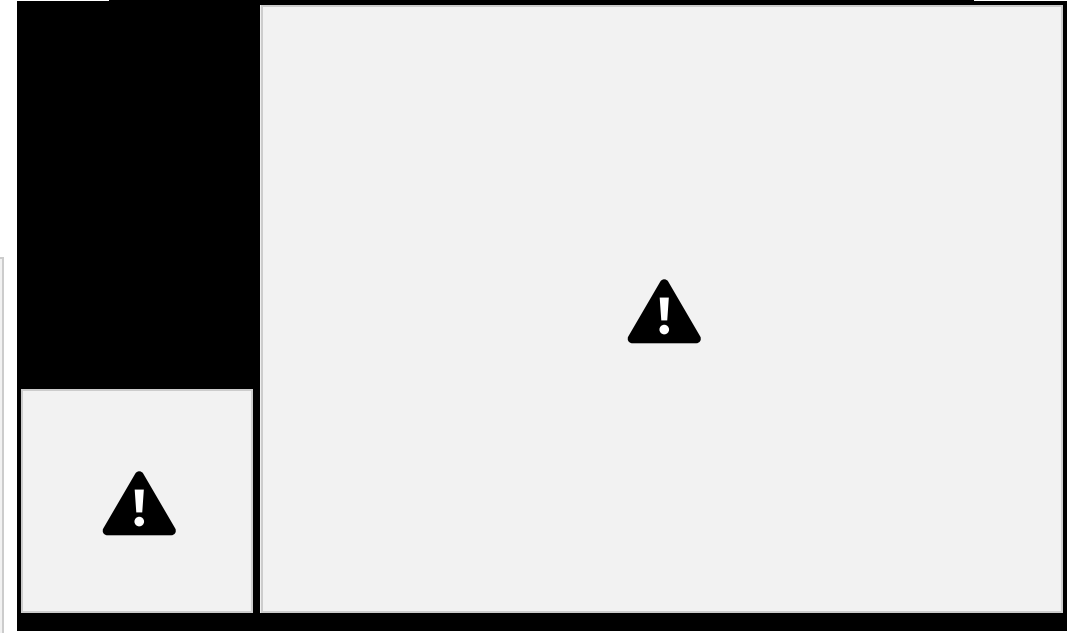
Market Basket Analysis





- En el análisis de asociación, nos interesa el número absoluto de transacciones realizadas que contienen un conjunto particular de elementos (items).
- Una aplicación típica del análisis de asociación es el análisis de la cesta de mercado, que se usa para observar el comportamiento de compra de los consumidores en los supermercados.
- Al encontrar relaciones entre los artículos que la gente compra, un minorista puede usar esta información para generar estrategias de marketing para incrementar sus ingresos.

Market Basket Analysis





El modelo de la cesta de mercado



Este modelo se basa en la relación entre los items y las transacciones, con algunas suposiciones sobre la forma de los datos.

- Cada transacción consta de un conjunto de items (itemset). La cantidad de productos en una transacción es menor que la cantidad total de items ofrecidos por un supermercado. □

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



El modelo de la cesta de mercado

El objetivo es identificar los itemsets frecuentes, es decir, los conjuntos de artículos que aparecen en muchas cestas.



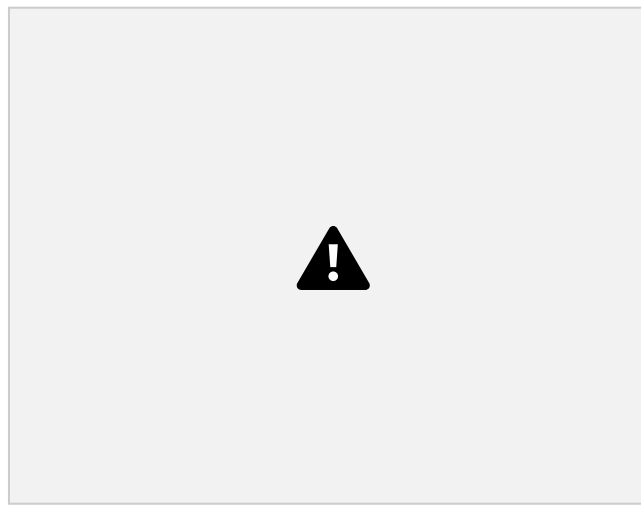


Figure 10: Base de datos de un supermercado

En nuestro ejemplo podemos ver que “manzana” se compra junto con “cerveza” en tres transacciones

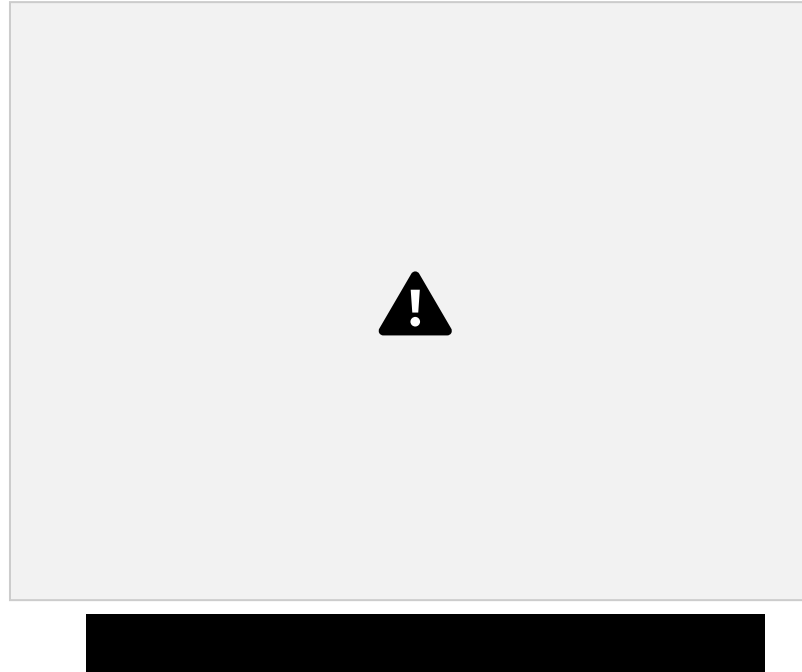
(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com



Reglas de asociación

Aunque estamos interesados en extraer conjuntos frecuentes de elementos, esta información a menudo se presenta como una colección de implicaciones del tipo “si

esto, entonces aquello”, denominadas reglas de asociación.



(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

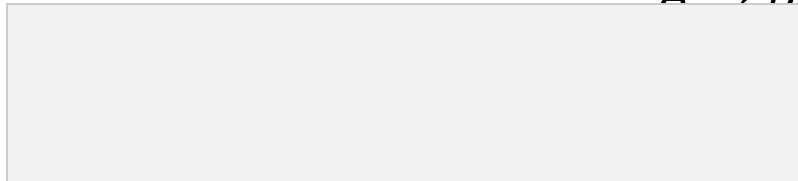
Reglas de asociación

La forma de una regla de asociación se ve como



$A \rightarrow B$

donde A es



un conjunto de elementos, también

llamado antecedente o lado izquierdo (LHS), y B es solo un elemento, también llamado consecuente o lado derecho (RHS). La regla se lee de la siguiente manera si todos los elementos de A aparecen en alguna cesta, entonces es “probable” que B también aparezca en esa cesta. Hay que notar que el símbolo de implicación \rightarrow no indica una relación causal entre A & B sino que es una estimación de la probabilidad condicional de B dado A . Un ejemplo de una regla de asociación podría ser

$$\{ \text{café, azúcar} \} \rightarrow \{ \text{leche} \}$$

lo que significa que si se compran café y azúcar, es probable que también se compre leche.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Medidas de Asociación

- Consideremos que tenemos una tienda de comestibles con decenas de miles de productos diferentes. No sería eficiente encontrar todas las reglas de asociación para todas



las posibles combinaciones de productos.

- En su lugar, querríamos seleccionar solamente las reglas potencialmente “relevantes”. Por lo tanto, utilizaremos medidas de asociación (soporte, confianza y elevación) para reducir el número de relaciones a analizar.

1. Soporte

Mide la frecuencia con la que se encuentra un itemset en todo el conjunto de transacciones. Puede calcularse como

$$\frac{\text{Número de transacciones que contienen el itemset}}{n}$$

Por lo general, se utiliza un umbral de soporte específico para reducir la cantidad de itemsets que necesitamos analizar. Al comienzo del análisis, podríamos establecer nuestro umbral de soporte entre el 1% y el 10%.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Medidas de Asociación





2. Confianza

Nos dice la probabilidad de que se compre B dado que se compro A , expresado como $A \rightarrow B$. Se calcula como:

$$\frac{\text{Soporte}(A \cup B)}{\text{Soporte}(A)}$$

Por tanto, la confianza se puede interpretar como una estimación de la probabilidad $P(B|A)$. La confianza

está dirigida, por lo que, generalmente da valores diferentes para las reglas $A \rightarrow B$ & $B \rightarrow A$.

Medidas de Asociación



3. Lift



Nos dice qué tan probable es que se compre B cuando se compra A mientras se consideran ambas popularidades, es decir, mide cuántas veces más A & B ocurren juntas si fueran estadísticamente independientes. La fórmula es

$$Lift(A \rightarrow B) = \frac{Confianza(A \rightarrow B)}{Soporte(A) \cdot Soporte(B)} = \frac{Soporte(A \& B)}{Soporte(A) \cdot Soporte(B)}$$

El valor de lift se puede interpretar como:

- $Lift = 1$: implica que no hay asociación entre A & B

- $Lift > 1$: implica que es probable que se compre B si se compra A



- $Lift < 1$: implica que es poco probable que se compre B si se compra

- A

Algoritmo A-Priori



- Existen varios algoritmos para el análisis de las reglas de asociación



(AIS, SETM, FP-Growth, Eclat), pero nosotros nos centraremos en

el algoritmo A-Priori que funciona eliminando conjuntos de elementos mirando primero los conjuntos más pequeños y reconociendo que un conjunto grande no puede ser frecuente a menos que todos sus subconjuntos lo sean.

En pocas palabras, el algoritmo establece que si un conjunto de elementos es poco frecuente, todos sus subconjuntos también deben serlo.

El algoritmo A-Priori utiliza una estrategia de búsqueda denominada en amplitud.

Con este principio, se pueden eliminar (podar) secciones del árbol de decisiones para limitar el número de conjuntos de elementos que deben examinarse.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Pasos del Algoritmo A-Priori

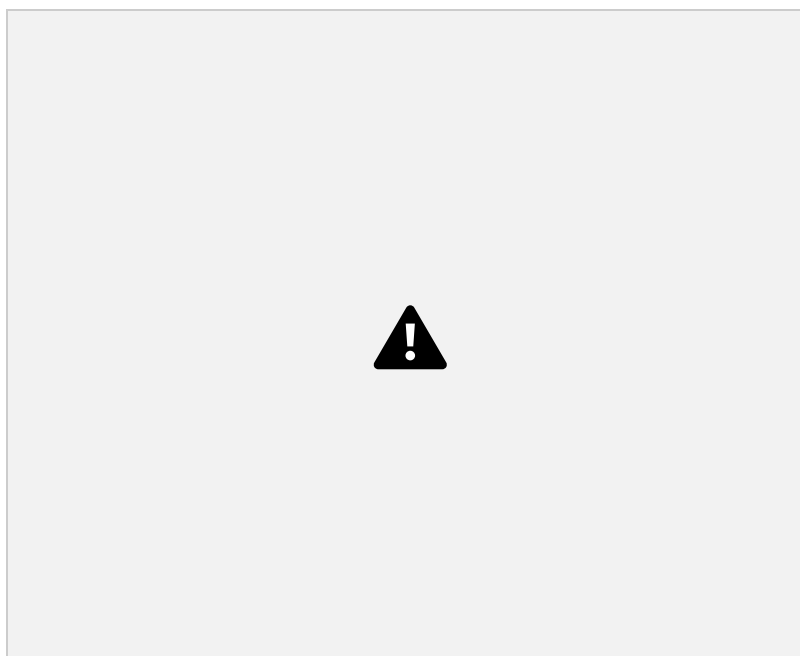




- Paso 1. Establecer un valor mínimo para el soporte y la confianza. Esto significa que solo nos interesa encontrar reglas para los elementos que tienen cierta existencia por defecto, por ejemplo, el apoyo, y tienen un valor mínimo de co-ocurrencia con otros elementos, por ejemplo, la confianza.
- Paso 2. Extraer todos los subconjuntos que tengan un valor de soporte superior a un umbral mínimo.
- Paso 3. Seleccionar todas las reglas de los subconjuntos con un valor de confianza superior al umbral mínimo.
- Paso 4. Ordenar las reglas por orden descendente de lift.



¿Que es RFM (Recency, Frecuency y Money)?



más de los

- El análisis de RFM es una técnica que se utiliza para agrupar o segmentar a los clientes existentes en función del comportamiento histórico con la esperanza de que la historia, con los motivadores adecuados, se repita o incluso mejore.

- El acrónimo es la abreviatura de actualidad, frecuencia y valor monetario y cada una de estas medidas se alinea con uno o

tres métodos para aumentar los ingresos de una

empresa.

- El análisis RFM es un método de segmentación de clientes que lo ayuda a orientar a sus clientes con el mensaje correcto en el lugar correcto en el momento correcto en función de tres puntos de datos clave.

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

¿Que es RFM (Recency, Frecuency y Money)?

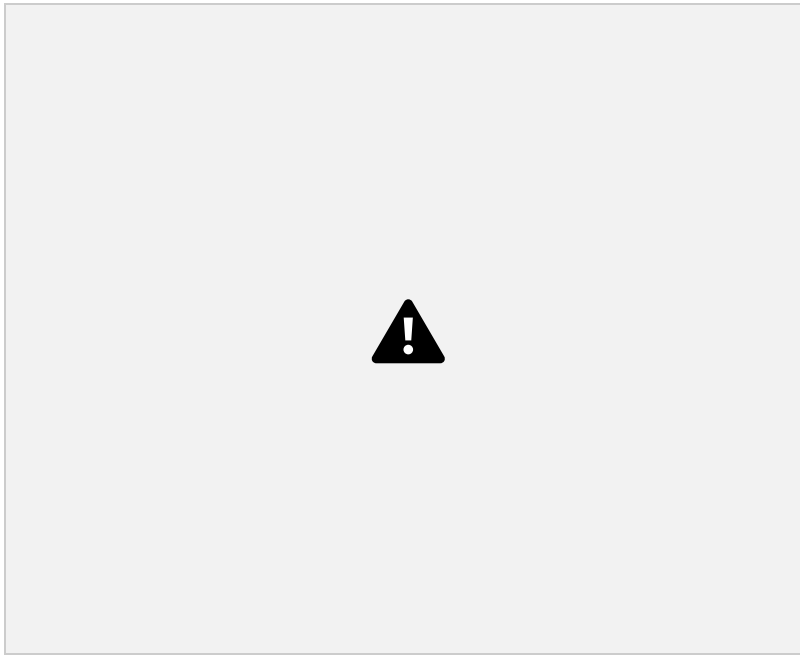




(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com

Beneficios del Análisis RFM





Realizar un análisis RFM en tu base de clientes y enviar campañas personalizadas a objetivos de alto valor tiene enormes beneficios para tu tienda de comercio electrónico.

- ❑ Personaliza: Creando segmentos de clientes efectivos, puedes crear ofertas relevantes y personalizadas.
- ❑ Mejora la tasa de conversión: Las ofertas personalizadas producen tasas de conversión más altas porque tus

clientes se comprometen con los productos que les interesan.

- ❑ Mejora la Economía Unitaria.
- ❑ Aumenta los ingresos y los beneficios.



Referencias Bibliográficas



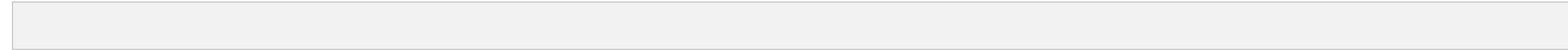
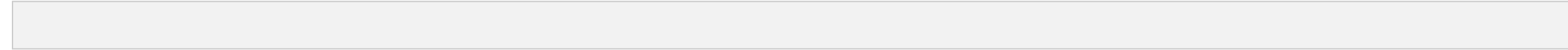
- An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics).
- Introduction to Machine Learning with Python: A Guide for Data Scientists.
- An Introduction to Statistical Learning by James, Gareth et al.
- Applied Predictive Modeling by Max Kuhn and

Kjell Johnson.

- <https://www.aprendemachinelearning.com/aplicaciones-del-machine-learning/>
- <https://aprendeia.com/todo-sobre-aprendizaje-supervisado-en-machine-learning/>

(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com





(+51) 976 760 803 www.datayanalytics.com info@datayanalytics.com