1. Explica qué son las reglas de asociación y proporciona algunos ejemplos de aplicación  
   concretos, tanto en el contexto médico como otro más general. Los ejemplos deben  
   estar apropiadamente referenciados.

Las **reglas de asociación** es una técnica de minería de datos empleada para encontrar patrones frecuentes en un conjunto de datos.

Para IBM, quien ofrece una definición práctica al definir que las **reglas de asociación** relacionan una determinada conclusión (por ejemplo, la compra de un producto dado) con un conjunto de condiciones (por ejemplo, la compra de otros productos). (IBM Documentación, 2020).

Una definición formal se basa en el problema de extraer reglas de asociación, entre conjuntos de ítems en una base de datos de transacciones de los clientes. Eventualmente, cada transacción consiste de ítems o atributos comprados por un cliente en una visita al comercio. Por tanto, el término transacción hace referencia a cada grupo de eventos que están asociados de alguna forma, por ejemplo: (Amat Rodrigo, 2018)

* La cesta de compras en un supermercado.
* Los libros que un cliente compra en una librería.
* Las páginas web que visita un usuario.

El algoritmo ARM (Association Rules Mining, por sus siglas en inglés) surge por el interés de encontrar un tipo de reglas que tienen (Agrawal, Imielinski, & Swami, 1993):

* Mínimo soporte transaccional *s* - la unión de ítems en el consecuente y antecedente de la regla está presente en un mínimo de *s%* transacciones en la base de datos.
* Mínima confidencia *c* – por lo menos *c%* de transacciones en la base de datos que satisface el antecedente de la regla, también satisface el consecuente de la regla.

Un ejemplo de las reglas de asociación, considerando el modelo de un supermercado para la definición del algoritmo por parte de Rakesh Agrawal, Tomasz Imielinski y Arun Swami (Agrawal, Imielinski, & Swami, 1993, pág. 207); es la mayoría de personas que compran pan, también compran leche. Por tanto, en lenguaje matemático sería:

La parte izquierda de la función se llama antecedente y la parte derecha se denomina consecuente (Jiménez Martínez, 2018). El descubrimiento de las reglas de asociación permite desarrollar estrategias de mercadeo con fines de aumentar las ventas, aunque también ha tenido aplicación en campos como la medicina, meteorología, detección de fraudes bancarios, prevención de fallas en aeronaves, entre otras áreas.

En cuanto a los estudios de mercadeo, se da el caso de una tienda de electrométricos, en donde los productos que los clientes compran con mayor frecuencia pueden ser ubicados en sitios distantes, mientras que los productos que se venden con mínima frecuencia se ubican en sitios cercanos a la vista de los clientes. Por ejemplo, si los clientes compran computadores y se inclinan por la compra de software financiero, entonces la ubicación del hardware puede ayudar a incrementar las ventas del software (De Moya Amaris & Rodríguez Rodríguez, 2003)

En el campo de la medicina, las reglas de asociación se han usado para identificar interacciones entre un conjunto de proteínas de VIH-1, y un conjunto de proteínas humanas con alta confidencia, dado que la identificación entre un huésped viral y una proteína – proteína tiene un enfoque importante y útil en el desarrollo de nuevos fármacos, permitiendo elegir el blanco de este tipo de interacciones (Mukhopadhyay, Maulik, Bandyopadhyay, & Eils, 2010).

El área de las imágenes médicas es adecuada para la extracción de información oculta o aún no descubierta, por medio del algoritmo de reglas de asociación. Imágenes de radiología (rayos X), así como imágenes digitales tales como Resonancia Magnética (MR, Magnetic Resonance), Ultrasonografía (US), Radiografía Computada Proyectada (CR, Projectional Computed Radiography), Tomografía Computada (CT, Computed Tomography) y Tomografía de Emisión de Posición (PET, Position Emission Tomography) se han usado ampliamente para tomar decisiones en cuanto al diagnóstico y pronóstico de una enfermedad. La naturaleza compleja de las imágenes médicas indica que al representar completamente dichas imágenes por medio de ARM, cada imagen o transacción tiene un amplio número de atributos o ítems (Olukunle & Ehikioya, 2002).

1. Caracterización del dataset: ¿Explica cómo has caracterizado los datos erróneos que deben limpiarse antes de seguir con el análisis y por qué crees que son erróneos? ¿Cuántos casos médicos y cuántos síntomas diferentes hay en el dataset? ¿Cuántos pacientes presentan uno o más síntomas? ¿Cuáles son los 5 síntomas más comunes?  
   Detalla todas las caracterizaciones descriptivas de este dataset que te parezcan  
   relevantes. (notast, 2019)

Con el fin de realizar de manera sencilla el tratamiento de los datos, se ha migrado el conjunto de datos a un formato CSV, sin alterar los registros originales del archivo. (Li, 2017)

define:association rules

<https://towardsdatascience.com/association-rules-2-aa9a77241654>

<https://www.solver.com/xlminer/help/association-rules>

<https://www.researchgate.net/publication/220907097_Basic_Association_Rules>

<https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html/2>

<https://www.kdnuggets.com/2016/03/datacamp-r-learning-path-7-steps.html>

<https://www.sciencedirect.com/topics/computer-science/association-rules>

define:reglas de asociación

<https://www.cienciadedatos.net/documentos/43_reglas_de_asociacion#Introducci%C3%B3n>

<https://rstudio-pubs-static.s3.amazonaws.com/367334_353f1bbf1b3543e180bb9210e711a73f.html#fnref4>

<https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=nodes-association-rules>

define: association rules in google scholar

<http://rakesh.agrawal-family.com/papers/vldb95tax_rj.pdf>

<https://ieeexplore.ieee.org/abstract/document/5735401>

<https://ieeexplore.ieee.org/abstract/document/1013116>

define:applications of association rules

<https://chat.whatsapp.com/HnICBSQbAmuBKD6hOyKyVP>

<https://t.me/joinchat/Z2CR6RPhmjZmMGIx>

<https://stackoverflow.com/questions/9126840/delete-rows-with-blank-values-in-one-particular-column>

<https://rpubs.com/camilamila/limpieza>

<https://rpubs.com/camilamila/limpieza_R>

<https://rpubs.com/bradleyboehmke/data_wrangling>

<https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>

<https://icd10cmtool.cdc.gov/?fy=FY2021&q=sarcopeni> Lista de enfermedades

<https://www.icd10data.com/ICD10CM/Codes/R00-R99/R25-R29>

<https://es.stackoverflow.com/questions/273034/estructurar-un-dataframe-para-arules>

<https://stackoverflow.com/questions/6386314/how-do-i-get-discrete-factor-levels-to-be-treated-as-continuous>

<https://dplyr.tidyverse.org/reference/count.html>

<https://stackoverflow.com/questions/13967063/remove-duplicated-rows>

<https://www.r-bloggers.com/2019/05/whats-that-disease-called-overview-of-icd-package/>

<https://rpubs.com/daniballari/ggplot>

<https://link.springer.com/article/10.1007/s00223-013-9758-y>

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, 215.

Amat Rodrigo, J. (Junio de 2018). *Ciencia de Datos*. Obtenido de https://www.cienciadedatos.net/documentos/43\_reglas\_de\_asociacion#Introducci%C3%B3n

De Moya Amaris, M. E., & Rodríguez Rodríguez, J. E. (2003). La contribución de las reglas de asociación a la minería de datos. *Tecnura*, 96. Obtenido de https://revistas.udistrital.edu.co/index.php/Tecnura/article/view/6175

Jiménez Martínez, C. (6 de 3 de 2018). *RStudio Pubs*. Obtenido de https://rstudio-pubs-static.s3.amazonaws.com/367334\_353f1bbf1b3543e180bb9210e711a73f.html#fnref4

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Eils, R. (2010). Mining Association Rules from HIV-Human Protein Interactions. *Proceedings of 2010 International Conference on Systems in Medicine and Biology*, 344-348. doi:10.1109/ICSMB.2010.5735401

Olukunle, A., & Ehikioya, S. (2002). A Fast Algorithm for Mining Association Rules in Medical Image Data. *Proceedingsof the 2002 IEEE Canadian Conference on Electrical & Computer Engineering*, 1181.