



Laboratorio 3 - Reglas de Asociación

Integrantes: Matías Escudero

Joaquín Macías

Curso: Análisis de Datos

Sección A-1

Profesor: Max Chacón Pacheco

Ayudante: Daniel Calderón

22 de Junio de 2023

Tabla de contenidos

1. Introducción	1
1.1. Objetivos	2
1.1.1. Objetivo General	2
1.1.2. Objetivos Específicos	2
2. Marco Teórico	4
2.1. Reglas de Asociación	4
2.2. Medidas de Calidad y Confianza	4
2.2.1. Soporte	4
2.2.2. Confianza	4
2.2.3. Lift	5
2.3. Monotonicidad	5
3. Obtención de Reglas	7
3.1. Pre-procesamiento de los datos	7
3.2. Generación de Reglas	8
3.3. Selección de Reglas	8
4. Análisis de Resultados y Comparación	10
4.1. TSH Alto \rightarrow Sin Hipertiroidismo	10
4.2. T4U Bajo, FTI Bajo \rightarrow TT4 Bajo	10
4.3. Mujer, T3 Alto, FTI Alto \rightarrow TT4 Alto	11
4.4. T3 Bajo, TT4 Bajo \rightarrow Sin Hipertiroidismo	11
4.5. TT4 Bajo \rightarrow Sin Hipertiroidismo	11
4.6. Mujer, TSH Alto \rightarrow Sin Hipertiroidismo	12
5. Conclusiones	13
Bibliografía	15

1. Introducción

Las reglas de asociación (Akash Saxena, 2021) son un tipo de técnica utilizada en el análisis de datos para descubrir patrones y relaciones interesantes entre los elementos de un conjunto de datos. Son particularmente útiles cuando se trata de descubrir relaciones entre diferentes items o características en un conjunto de datos.

En el contexto del Thyroid Disease Dataset, las reglas de asociación podrían ayudarnos a descubrir relaciones entre diferentes características o variables relacionadas con la enfermedad tiroidea. Estas reglas nos permiten identificar combinaciones de características o condiciones que ocurren con mayor frecuencia o que están asociadas entre sí.

Un ejemplo de una regla de asociación podría ser: *Si un paciente tiene un nivel anormalmente alto de TSH, es probable que también tenga un nivel anormalmente bajo de T4U*. Esta regla sugiere una asociación entre los niveles de TSH y T4U en los pacientes y puede ser útil para entender las interacciones entre estas variables y su impacto en el diagnóstico o tratamiento de la enfermedad tiroidea.

Para descubrir estas reglas de asociación, se utilizan algoritmos específicos, como el algoritmo Apriori (Xie, 2021), que busca patrones frecuentes en los datos y genera reglas basadas en su frecuencia y confianza. La frecuencia indica qué tan a menudo ocurre un patrón en el conjunto de datos, mientras que la confianza mide la probabilidad condicional de que ocurra un resultado dado un conjunto de condiciones.

Al descubrir y comprender las reglas de asociación en el conjunto de datos, podemos obtener información valiosa sobre las relaciones entre las variables y características asociadas con la enfermedad tiroidea. Estas reglas pueden ayudar a los médicos y profesionales de la salud a tomar decisiones más informadas sobre el diagnóstico, el tratamiento y la gestión de la enfermedad.

1.1. Objetivos

1.1.1. Objetivo General

- El objetivo principal de las reglas de asociación en el conjunto de datos es descubrir patrones interesantes y significativos entre las características y variables relacionadas con la enfermedad tiroidea. Estos patrones pueden ayudar a identificar relaciones ocultas y comprender mejor la enfermedad.
- Al descubrir reglas de asociación relevantes, se pueden identificar combinaciones de características o condiciones que están asociadas con ciertos diagnósticos o estados de la enfermedad tiroidea. Esto puede mejorar la precisión del diagnóstico y ayudar a los médicos a tomar decisiones más informadas.
- Las reglas de asociación pueden proporcionar información sobre las relaciones entre diferentes variables y cómo se relacionan con los resultados del tratamiento. Esto puede ayudar a desarrollar enfoques de tratamiento más personalizados y efectivos para los pacientes con enfermedad tiroidea.
- Al analizar las reglas de asociación, es posible identificar factores de riesgo o combinaciones de características que están asociadas con un mayor riesgo de desarrollar enfermedad tiroidea. Esto puede ayudar en la identificación temprana de individuos en riesgo y en la implementación de estrategias preventivas.

1.1.2. Objetivos Específicos

1. Descubrir patrones de asociación entre los niveles de hormonas tiroideas y las características demográficas, como la edad y el sexo.
2. Identificar combinaciones de síntomas y resultados de pruebas de laboratorio que están asociadas con diagnósticos específicos de enfermedad tiroidea.
3. Evaluar la asociación entre las características clínicas y los resultados de tratamiento en pacientes con enfermedad tiroidea.

4. Identificar combinaciones de factores de riesgo, como antecedentes familiares y exposiciones ambientales, que están asociadas con un mayor riesgo de desarrollar enfermedad tiroidea.
5. Evaluar la asociación entre los patrones de medicación y la respuesta al tratamiento en pacientes con enfermedad tiroidea.

2. Marco Teórico

2.1. Reglas de Asociación

Dentro del ámbito de la minería de datos las reglas de asociación son utilizadas para **descubrir hechos que ocurren en común dentro de un mismo conjunto de datos** (Menzies and Hu, 2003). La idea básica de reglas de asociación proviene del análisis de mercado de un carro de compras, donde se busca encontrar patrones relacionados a qué productos se compran juntos frecuentemente.

Una regla de asociación es una implicación de la forma $X \rightarrow Y$, donde X e Y son conjuntos de elementos. X e Y deben ser disjuntos, es decir, no compartir ningún elemento. X se conoce cómo el antecedente (o *left-hand-side*) de la regla e Y como el consecuente (o *right-hand-side*) (Agrawal et al., 1993).

2.2. Medidas de Calidad y Confianza

2.2.1. Soporte

El soporte de una regla de asociación es la proporción de transacciones o instancias en el conjunto de datos que contienen tanto el antecedente como el consecuente de la regla. Es decir, es la frecuencia relativa con la que ocurre la regla en el conjunto de datos. Un alto soporte indica que la regla es común y se encuentra en un número significativo de transacciones o instancias (Jose Diaz-Garcia, 2023).

2.2.2. Confianza

La confianza de una regla de asociación mide la probabilidad condicional de que el consecuente de la regla ocurra dado que el antecedente también ocurra. Es decir, es la proporción de transacciones o instancias que contienen tanto el antecedente como el consecuente en comparación con aquellas que solo contienen el antecedente. Una alta confianza indica que el consecuente es probablemente verdadero cuando se cumple el antecedente (Fuguang Bao, 2022).

2.2.3. Lift

El lift de una regla de asociación mide la relación entre la confianza observada de la regla y la confianza esperada si el antecedente y el consecuente fueran independientes. Un lift mayor a 1 indica que la ocurrencia del consecuente es más probable cuando se cumple el antecedente que si fueran eventos independientes. Un lift menor a 1 indica una asociación negativa, lo que sugiere que la ocurrencia del consecuente es menos probable cuando se cumple el antecedente (P.D. McNicholas, 2008).

Estas medidas proporcionan información sobre la calidad y la fuerza de las reglas de asociación encontradas en el conjunto de datos. El soporte indica cuán común es la regla en el conjunto de datos, la confianza mide la probabilidad condicional y el lift mide la relación entre la confianza observada y la esperada. Al utilizar estas medidas, se pueden identificar las reglas de asociación más relevantes y significativas en el conjunto de datos, lo que puede ayudar a comprender mejor las relaciones y patrones entre las variables relacionadas con la enfermedad tiroidea.

2.3. Monotonidad

La monotonidad es una propiedad que se puede aplicar a las reglas de asociación en el contexto de la minería de datos y el análisis de reglas. La monotonidad se refiere a la relación ordenada entre los elementos del antecedente y el consecuente de una regla de asociación (Dinesh J. Prajapati, 2017).

En el contexto de las reglas de asociación del conjunto de datos, la monotonidad se refiere a si la ocurrencia de un conjunto de características en el antecedente de la regla implica la ocurrencia de otro conjunto de características en el consecuente, o viceversa. En otras palabras, si se cumple una condición, entonces se cumple otra.

Por ejemplo, supongamos que se descubre la siguiente regla de asociación en el conjunto de datos: *Si un paciente tiene un nivel anormalmente alto de TSH, entonces es probable que también tenga un nivel anormalmente bajo de T₄U*. En este caso, se podría decir que la regla

muestra monotonicidad, ya que la ocurrencia de un nivel alto de TSH implica la ocurrencia de un nivel bajo de T4U.

La monotonicidad es una propiedad importante en las reglas de asociación, ya que permite establecer una relación lógica y ordenada entre las características en el antecedente y el consecuente. Ayuda a comprender cómo se relacionan las variables y cómo una característica puede influir en la presencia o ausencia de otra característica. Esto puede ser útil en la interpretación y aplicación de las reglas de asociación en el contexto del Thyroid Disease Dataset, ya que proporciona una comprensión más profunda de las interacciones entre las variables relacionadas con la enfermedad tiroidea.

3. Obtención de Reglas

3.1. Pre-procesamiento de los datos

Para la creación de reglas de asociación se busca encontrar relaciones entre distintas categorías del conjunto de datos con el cual se está trabajando. Para la creación y análisis de estas reglas se utiliza el conjunto de datos `allhpyer` que se ha utilizado en las experiencias anteriores, en este caso, con el set de datos ya limpio y con datos imputados de acuerdo a la experiencia anterior. Sin embargo, de acuerdo al análisis exploratorio realizado en el primer laboratorio se evidenció que existen distintas variables booleanas que presentan incluso más de un 99 % de sus valores FALSO. El hecho de trabajar con variables tan desequilibradas puede generar una gran cantidad de reglas de asociación que no entreguen ninguna información relevante, por lo cual **se decide eliminar estas columnas con muy baja variabilidad** (*query on thyroxine, on antithyroid medication, pregnant, thyroid surgery, I131 treatment, lithium, goitre, tumor y hypopituitary*) para disminuir el ruido en los resultados finales.

Además, al estar trabajando con reglas de asociación es importante reclasificar las variables sobre las cuales se trabaja para así facilitar la interpretación de las reglas a futuro, para lo cual **todas las variables se transforman en variables categóricas**. Si bien muchas variables del conjunto de datos son categóricas, importa reclasificar las variables numéricas, para las cuales se siguieron los siguientes criterios:

- La edad se separa en tres grupo etarios: **Adultos jóvenes** (Hasta 29 años), **Adultos** (entre 30 y 59 años) y **Mayores de Edad** (más de 60 años). Esta decisión fue tomada en base a literatura que sustenta que estos rangos se utilizan para estudiar casos de hipertiroidismo teniendo en cuenta el grupo etario del paciente (Vanderpump (2011)).
- Las variables clínicas (TSH, T3, TT4, FTI Y T4U) se categorizan en base a los valores que presenta cada una de estas variables, **clasificándolas en nivel bajo, medio-bajo, medio-alto y alto de acuerdo a sus respectivos rangos inter-cuartílicos**.

3.2. Generación de Reglas

Ya con el conjunto de datos con todas sus variables clasificadas, se procede a generar reglas de asociación. Para lograr esto se utiliza el algoritmo `a priori` del paquete `arules` de R, que busca encontrar reglas que expliquen con precisión cómo se asocian distintos atributos en distintas transacciones. Se entregan dos parámetros a esta función: **soporte mínimo** y **confianza mínima**. Se indica cómo soporte mínimo 0.05, lo cual significa que interesa que algoritmo calcule reglas que se apliquen al menos al 5 % de las transacciones. Por otra parte, se indica una confianza mínima de 0.7, que significa que se buscan reglas que sean verdaderas al menos el 70 % de las veces. Recordar que estos valores son preliminares y luego se aplicarán filtros por sobre las reglas que sean generadas a partir de los parámetros previamente indicados. A partir de la ejecución del algoritmo se obtuvieron 461 reglas, las cuales serán filtradas en la siguiente sección.

3.3. Selección de Reglas

A continuación, de las 461 reglas generadas se muestran las 10 primeras, ordenadas según su confianza, lift y soporte:

LHS	RHS	Support	Confidence	Coverage	Lift	Count
TT4=Medio-Bajo	hyperthyroid=Sin Hipertiroidismo	0.24606017	1	0.24606017	1.025716	687
TSH=Alto	hyperthyroid=Sin Hipertiroidismo	0.24462751	1	0.24462751	1.025716	683
TSH=Medio-Alto	hyperthyroid=Sin Hipertiroidismo	0.23674785	1	0.23674785	1.025716	661
sex=Mujer, FTI=Bajo	hyperthyroid=Sin Hipertiroidismo	0.18588825	1	0.18588825	1.025716	519
sex=Mujer, TSH=Alto	hyperthyroid=Sin Hipertiroidismo	0.18230659	1	0.18230659	1.025716	509
sex=Mujer, TSH=Medio-Bajo	hyperthyroid=Sin Hipertiroidismo	0.17227794	1	0.17227794	1.025716	481
sex=Mujer, TT4=Medio-Bajo	hyperthyroid=Sin Hipertiroidismo	0.16081662	1	0.16081662	1.025716	449
sex=Mujer, TT4=Bajo	hyperthyroid=Sin Hipertiroidismo	0.15866762	1	0.15866762	1.025716	443
sex=Mujer, TSH=Medio-Alto	hyperthyroid=Sin Hipertiroidismo	0.15759312	1	0.15759312	1.025716	440
T3=Bajo, TT4=Bajo	hyperthyroid=Sin Hipertiroidismo	0.12786533	1	0.12786533	1.025716	357

Cuadro 1: 10 reglas de asociación con mayor soporte, lift y confianza

A partir de la tabla anterior es posible notar que para todos los casos con mejores índices de soporte, confianza y lift el consecuente (RHS) en todos los casos son casos negativos por hipertiroidismo. Si bien es interesante estudiar estos casos, ya que son reglas con los

mejores índices de calidad y confianza, era esperable tener resultados de este tipo debido a lo desbalanceado de esta variable (más del 97 % casos negativos). Para añadir reglas que no sean únicamente para los casos negativos por hipertiroidismo, se muestran las 10 reglas con mayores índices de soporte, lift y confianza que no contengan casos negativos de hipertensión como consecuente:

LHS	RHS	Support	Confidence	Coverage	Lift	Count
T4U=Bajo, FTI=Bajo	TT4=Bajo	0.05551576	0.9748428	0.05694842	3.769752	155
T4U=Bajo, FTI=Bajo, hyperthiroid=Sin Hipertiroidismo	TT4=Bajo	0.05551576	0.9748428	0.05694842	3.769752	155
sex=Mujer, T3=Alto, FTI=Alto	TT4=Alto	0.05193410	0.9294872	0.05587393	3.856060	145
TSH=Alto, T4U=Alto	sex=Mujer	0.05945559	0.8972973	0.06626074	1.296715	166
TSH=Alto, T4U=Alto, hyperthiroid=Sin Hipertiroidismo	sex=Mujer	0.05945559	0.8972973	0.06626074	1.296715	166
T3=Alto, FTI=Alto	TT4=Alto	0.06375358	0.8900000	0.07163324	3.692244	178
TSH=Bajo, T3=Alto, TT4=Alto	sex=Mujer	0.05265043	0.8698225	0.06053009	1.257011	147
T4U=Alto, FTI=Medio-Bajo	sex=Mujer	0.05050143	0.8597561	0.05873926	1.242463	141
T4U=Alto, FTI=Medio-Bajo, hyperthiroid=Sin Hipertiroidismo	sex=Mujer	0.05014327	0.8588957	0.05838109	1.241220	140
TT4=Alto, T4U=Alto	sex=Mujer	0.09240688	0.8514851	0.10852436	1.230511	258

Cuadro 2: Top 10 reglas sin casos sin hipertiroidismo en RHS.

A partir de las tablas anteriores es posible notar que muchas reglas generadas se relacionan al sexo femenino, y niveles de TSH, TT4, T3 y FTI, que son variables importantes dentro del conjunto de datos trabajo de acuerdo a las experiencias de laboratorio anteriores. A partir de las tablas se extraen reglas que parecen interesantes, las cuales serán analizadas en la siguiente sección.

4. Análisis de Resultados y Comparación

A continuación, se extraen y analizan reglas de asociación recabadas a partir de las tablas mostradas anteriormente y que consideramos importantes en base a las experiencias de laboratorio realizadas anteriormente y según sustento bibliográfico.

4.1. TSH Alto \rightarrow Sin Hipertiroidismo

Esta regla es congruente de acuerdo a lo visto en las experiencias anteriores y a los conocimientos médicos actuales. Es sabido que la Hormona Estimulante de la Tiroides (TSH) es producida por la glándula pituitaria y estimula a la glándula tiroides para que produzca más hormonas tiroideas. En el caso del hipertiroidismo, donde los niveles de hormonas tiroideas son elevados, los niveles de TSH disminuyen como parte de un mecanismo de retroalimentación negativa (of Diabetes et al. (2021)). Dicho esto, **los niveles altos de TSH se asocian a hipotiroidismo, no hipertiroidismo**. Esta regla tiene una alta confianza (1.0) y un lift mayor a 1 (1.03), lo que indica una fuerte asociación entre estos elementos. Los resultados obtenidos por la regresión logística de la primera experiencia de laboratorio indicaron que una disminución de los niveles de TSH aumentaban la probabilidad de padecer hipertiroidismo, mientras que en la segunda experiencia esta variable no fue particularmente determinante sobre la forma en que se realizó el agrupamiento.

4.2. T4U Bajo, FTI Bajo \rightarrow TT4 Bajo

Esta regla sugiere que cuando los niveles de T4U (Tiroxina sin unir) y FTI (Índice de Tiroxina Libre) son bajos, es probable que los niveles de TT4 (Tiroxina Total) también sean bajos. Al igual que la regla anterior, esta regla es congruente con la literatura médica existente, ya que T4U y FTI son medidas relacionadas con la cantidad de tiroxina en la sangre (Midgley (2001)). Esta regla presenta una confianza muy alta (0.97) y un lift de 3.77, lo que sugiere una asociación muy fuerte entre estas dos variables. Respecto a experiencias anteriores, mediante la matriz de correlación de las variables se vio correlación positiva entre estas tres variables, además, en la segunda experiencia se generó un grupo en donde estas tres variables eran altas simultáneamente.

4.3. Mujer, T3 Alto, FTI Alto \rightarrow TT4 Alto

La regla indica que para mujeres con niveles altos de T3 (triyodotironina) y FTI, es probable que los niveles de TT4 también sean altos. Similar a la regla anterior, hace sentido desde un punto de vista médico, ya que tanto T3 como FTI son indicadores de la actividad tiroidea, y los niveles elevados de estos generalmente están asociados con una mayor producción de hormonas tiroideas (Midgley (2001)). Además, las condiciones médicas asociadas a la tiroides tienden a ser más comunes en las mujeres que en los hombres (Li and Li (2015)), lo que puede explicar la presencia del sexo femenino en esta regla. Esta regla tiene una confianza bastante alta (0.9294872) y un lift de 3.856060, lo que implica una fuerte asociación entre estos elementos. Además, hace sentido de acuerdo a las dos experiencias anteriores, donde el sexo femenino se vio reflejado cómo una variable importante dentro del fenómeno estudiado.

4.4. T3 Bajo, TT4 Bajo \rightarrow Sin Hipertiroidismo

Esta regla sugiere que cuando los niveles de T3 y TT4 son bajos, es probable que no haya hipertiroidismo. También tiene sentido acorde a la literatura médica, ya que los niveles bajos de estas hormonas tiroideas suelen estar asociados con el hipotiroidismo, no con el hipertiroidismo Fisher (1996). La regla tiene una confianza perfecta (1.0) y un lift de 1.03, lo que indica una fuerte asociación entre estos elementos. Similar a las reglas anteriores, las experiencias anteriores indican que el fenómeno de hipertiroidismo está ligado a un aumento en estas hormonas.

4.5. TT4 Bajo \rightarrow Sin Hipertiroidismo

Muy similar a la regla anterior, esta regla sugiere que los bajos niveles de TT4 indican la ausencia de hipertiroidismo, que se alinea con la literatura médica existente ya vista. La regla tiene una confianza perfecta (1.0) y un lift de 1.03, lo que indica una fuerte asociación entre estos elementos.

4.6. Mujer, TSH Alto \rightarrow Sin Hipertiroidismo

Esta regla indica que para mujeres con niveles altos de TSH es probable que no haya hipertiroidismo. Como se mencionó anteriormente, los altos niveles de TSH generalmente indican hipotiroidismo. Si bien en una de las reglas mencionadas anteriormente el sexo femenino se asocia a un aumento en la probabilidad de sufrir hipertiroidismo, es sabido que las mujeres tienen un mayor riesgo de trastornos de la tiroides, incluyendo el hipotiroidismo (Li and Li (2015)). Esta regla tiene una confianza perfecta (1.0) y un lift de 1.03, lo que sugiere una fuerte asociación entre estos elementos.

5. Conclusiones

En el presente informe se lograron generar e identificar distintas reglas de asociación para descubrir relaciones interesantes entre varios atributos relacionados con la tiroides. Estas relaciones no sólo han demostrado ser estadísticamente significativas, sino que también han sido corroboradas por la literatura médica existente.

Aterrizado al contexto del problema con el cuál se está trabajando, las reglas de asociación también pueden revelar combinaciones de factores de riesgo o características que están asociados con un mayor riesgo de desarrollar enfermedad tiroidea. Por ejemplo, se podría encontrar una regla que establezca que *si un paciente tiene antecedentes familiares de enfermedad tiroidea y ha estado expuesto a ciertos factores ambientales, es más probable que desarrolle hipotiroidismo*. **Esto puede ayudar en la identificación de grupos de riesgo y en la implementación de medidas preventivas.**

Además, las reglas de asociación pueden proporcionar información sobre cómo ciertas características o variables están asociadas con la respuesta al tratamiento en pacientes con enfermedad tiroidea. Por ejemplo, se podría encontrar una regla que establezca que *si un paciente tiene ciertos valores de hormonas tiroideas antes del tratamiento, es más probable que responda positivamente al tratamiento con medicamentos específicos*. Esto puede ayudar a personalizar el tratamiento y mejorar los resultados para los pacientes.

Sumado a lo anterior, las reglas de asociación también pueden revelar interacciones interesantes y complejas entre las variables relacionadas con la enfermedad tiroidea. Por ejemplo, se podría encontrar una regla que establezca que *si un paciente tiene niveles anormales de hormonas tiroideas y también tiene ciertas características demográficas, es más probable que desarrolle complicaciones adicionales*. Esto puede ayudar a comprender mejor las interacciones entre las variables y su impacto en la enfermedad.

Respecto a la etapa previa a la generación de las reglas de asociación, se generó una reclasificación de las variables. Es importante notar que la forma en la cual se realiza esta reclasificación puede afectar en las reglas generadas. El análisis de reglas de asociación debido a los cambios en la distribución de los datos y la representación de las relaciones entre las variables. La

reclasificación implica agrupar los valores originales de una variable en categorías diferentes o cambiar la escala de los valores.

Además, se evidenció como las variables desbalanceadas en el conjunto de datos pueden generar mucho ruido en el análisis de reglas de asociación debido a la falta de representatividad de ciertos grupos o categorías en los datos. De acuerdo a esto, **se hace muy importante limpiar correctamente el conjunto de datos antes de construir las reglas de asociación**, en caso contrario se pueden generar muchas reglas que no concluyen nada relevante.

Finalmente, mencionar que los métodos de reglas de asociación, regresión logística y k-means tienen sus propias fortalezas y debilidades, y **la elección del método a utilizar va a depender del problema y los datos en cuestión**. En el contexto de este estudio, las reglas de asociación ofrecen un enfoque intuitivo para descubrir relaciones ocultas entre las características, aunque a veces pueden generar una gran cantidad de reglas a analizar. En comparación, la regresión logística es útil cuando buscamos un modelo predictivo que estime la probabilidad de un evento; sin embargo, necesita un preprocesamiento cuidadoso de los datos y puede ser sensible a los valores atípicos. Por último, el método k-means es muy útil cuando se quiere dividir un conjunto de datos en grupos distintos basados en similitudes.

Bibliografía

- Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pages 207–216.
- Akash Saxena, V. R. (2021). A Comparative Analysis of Association Rule Mining Algorithms.
- Dinesh J. Prajapati, S. G. (2017). Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment.
- Fisher, D. A. (1996). Physiological variations in thyroid hormones: physiological and pathophysiological considerations. *Clinical Chemistry*, 42(1):135–139.
- Fuguang Bao, L. M. (2022). An Improved Evaluation Methodology for Mining Association Rules.
- Jose Diaz-Garcia, M. D. R. (2023). A survey on the use of association rules mining techniques in textual social media.
- Li, H. and Li, J. (2015). Thyroid disorders in women. *Minerva medica*, 106(2):109–114.
- Menzies, T. and Hu, Y. (2003). Data mining for very busy people. *Computer*, 36(11):22–29.
- Midgley, J. E. (2001). Direct and indirect free thyroxine assay methods: theory and practice. *Clinical Chemistry*, 47(8):1353–1363.
- of Diabetes, N. I., Digestive, and Diseases, K. (2021). Thyroid tests. National Institute of Diabetes and Digestive and Kidney Diseases.
- P.D. McNicholas, T. M. (2008). Standardising the lift of an association rule.
- Vanderpump, M. P. (2011). The epidemiology of thyroid disease. *British medical bulletin*, 99(1).
- Xie, H. (2021). Research and Case Analysis of Apriori Algorithm Based on Mining Frequent Item-Sets.