

UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INFORMÁTICA



Laboratorio 3 - Reglas de Asociación

Integrantes: Gonzalo Cuevas

Sofía Castro

Curso: Análisis de Datos

Sección A-1

Profesor: Max Chacón Pacheco

7 de Septiembre de 2022

Tabla de contenidos

1. Introducción	1
1.1. Objetivos	1
2. Marco Teórico	2
2.1. Reglas de Asociación	2
2.2. Medidas de Calidad y Confianza	3
2.2.1. Propiedades de las Medidas	3
2.3. Algoritmo A-Priori	4
3. Resultados	6
3.1. Pre-procesamiento	6
3.2. Soporte de las Variables	6
3.3. Soporte de la Reglas	8
3.4. Reglas Escogidas	9
4. Análisis de Resultados	12
4.1. Consecuente según clase	12
4.2. Consecuente según hormona	13
4.3. Consecuente embarazadas	14
4.4. Comparación con agrupamientos	15
5. Conclusión	16
Bibliografía	17

Índice de cuadros

1.	Soportes principales.	7
2.	Soportes para las clases.	7
3.	Soportes para las hormonas en su rango normal.	8
4.	Soportes para las reglas.	8
5.	Reglas escogidas.	11

Índice de figuras

1.	Reglas de asociación; consecuente = hombre	9
2.	Reglas de asociación; consecuente = TT4 bajo	9

1. Introducción

En el presente documento se lleva a cabo el desarrollo del laboratorio 3, el cual consiste en obtener reglas de asociación entre las variables pertenecientes a la base de datos *allhypo*.

Según Kotsiantis and Kanellopoulos (2006), las reglas de asociación corresponde a una técnica de la minería de datos que es capaz de reconocer patrones ocurrentes entre las variables. Su uso recae en diversas áreas, tales como, redes de telecomunicaciones, la gestión de mercados y riesgos, el control de inventario, etc (Kotsiantis and Kanellopoulos (2006)).

Sin embargo, el proceso implementado tiene como objetivo encontrar patrones escondidos y/o patrones que puedan ser corroborados con los resultados obtenidos en el laboratorio anterior.

Este informe consta de un marco teórico que permite comprender y obtener el conocimiento básico, para entender la siguiente sección que corresponde a el desarrollo, donde se exponen los procedimientos realizados en la base de datos y la implementación de las reglas de asociación, obteniendo los resultados. Luego en el apartado de análisis de resultados se genera una evaluación de las reglas obtenidas y una comparación con los resultados del laboratorio 2. Finalmente, se presentan las conclusiones del trabajo, conteniendo tanto las dificultades como los éxitos que conlleva.

1.1. Objetivos

1. Utilizar el paquete *arulesViz*.
2. Extraer conocimiento del problema asignado, por medio de las reglas de asociación.

2. Marco Teórico

En la siguiente sección se explican las diversas medidas que se han implementado para llevar a cabo el laboratorio.

2.1. Reglas de Asociación

En las técnicas de minería de datos se utilizan comúnmente patrones conocidos como reglas de asociación, las cuales intuitivamente reconocen patrones ocurrentes. La ventaja de su aplicación corresponde a facilitar la interpretación y visualización de los resultados generados (Lent et al. (NA)). Se basa en la expresión estipulada en la ecuación 1, donde X e Y son sets de atributos, su significancia recae en los valores de la base de datos, con X equivalente a True y la tendencia de Y igual a True, esto se traduce a que los atributos que poseen X tienden a contener Y a la vez (Toivonen et al. (1995)). El problema de esta técnica consiste en encontrar todas las reglas con un soporte y confianza mínima que satisfaga a un usuario en específico, estas medidas se explican en la siguiente sub-sección (Srikant and Agrawal (1995)).

$$X \Rightarrow Y \quad (1)$$

Existen tres tipos de reglas, que son las siguientes:

1. Conjuntiva: En cuyo caso $X \subseteq V$ será verdadero, si y solo si, todas las condiciones de X son verdaderas: $V_1 \wedge V_2 \wedge \dots \wedge V_m$ (Chacón (2015)).
2. Disyuntiva: En cuyo caso $X \subseteq V$ será verdadero, si y solo si, una o mas condiciones de X son verdaderas: $V_1 \vee V_2 \vee \dots \vee V_m$ (Chacón (2015)).
3. Interesantes: Cumple con un soporte mínimo y una confianza mínima, pudiendo ser una función entre ellas (Chacón (2015)).

Donde:

$$V = \bigcup_{q \in Q} V_i^q \quad (2)$$

Con V_i^q equivalente al dominio e indica los posibles valores de cada atributo o instancia y Q un conjunto finito, no vacío, de p atributos q_1, q_2, \dots, q_p , llamados también productos o ítems.

2.2. Medidas de Calidad y Confianza

El soporte y la confianza son dos medidas que reflejan la utilidad y la certeza de una regla de asociación, por parte del soporte, este se encuentra ligado al porcentaje de todas las transacciones que cumple en tener ambos conjuntos de una regla, mientras que la confianza es la probabilidad de que una transacción que contiene un conjunto de la regla, también contenga el otro. Amaris et al. (2003)

Matemáticamente, el soporte y confianza para la regla definida en la ecuación 1.

$$soporte(X \Rightarrow Y) = P(X \cup Y) \quad (3)$$

$$confianza(X \Rightarrow Y) = P(Y|X) = \frac{soporte(X \cup Y)}{soporte(X)} \quad (4)$$

Con esto, es de interés la búsqueda de reglas que abarquen un gran porcentaje de los datos, es decir, que cumplan con el mínimo soporte indicado, las cuales son llamadas reglas frecuentes, así como también, una regla es considerada apropiada cuando cumple con el mínimo soporte y la mínima confianza, es decir, posee una gran cobertura y precisión.

2.2.1. Propiedades de las Medidas

Las propiedades estipuladas a continuación consideran el tratamiento de reglas conjuntivas.

1. Especialización: La búsqueda de los posibles conjuntos X se realiza agregando condiciones al conjunto (Chacón (2015)).
2. Generalización: Ocurre cuando el antecedente de la regla contiene menos condiciones (Chacón (2015)).
3. Monotonía: Este criterio condiciona las medidas mencionadas, pues se asocia con el cumplimiento de las restricciones en función de la especialización. Se dice que la

medida o restricción es monótona si: $med(X_1) \leq med(X_2)$ y es anti-monótona si: $med(X_1) \geq med(X_2)$, palabras simples, para que sea monótona debe haber un crecimiento o decrecimiento constante. El soporte es monótono, puesto que la especialización de la regla lleva a mantener o disminuir el soporte, mientras que la confianza es anti-monótona (Chacón (2015)).

2.3. Algoritmo A-Priori

Permite generar todas las posibles combinaciones para un antecedente X, y durante la generación de las reglas, interrumpir la especialización usando medidas monótonas para una pre-poda, se utiliza esta combinatoria para generar un árbol paso a paso y realizando una pre-poda por soporte. Cada nodo g en el árbol es representado por dos grupos, uno es llamado cabeza h(g) que representa la regla para el nodo y el otro se llama cola t(g) que representa todas las posibles combinaciones (ordenadas) que pueden ser agregadas a la cabeza para formar una regla. Cuando el algoritmo encuentra un nodo h(g) cuyo soporte es inferior al minsop no genera las combinaciones, puesto que todas ellas no cumplirán con minsop. Por otro lado, ordenar la totalidad de las reglas generadas y entrega las confianzas que cumplen con minconf (Chacón (2015)). Para evaluar la confianza, se han generado diferentes medidas o métricas, las cuales exigen que sean monótonas en confianza, o soporte y confianza, pero manteniendo uno de ellas constante, estas medidas son explicadas a continuación:

1. Lift: Representa (ecuación 5) una medida de independencia entre X e Y, por lo que, su valor más bajo (1) es cuando X e Y sean completamente independientes (Chacón (2015)).

$$lift = \frac{nconf(X \Rightarrow Y)}{sop(Y)} \quad (5)$$

2. Convicción: También mantiene la monotonidad de la confianza, al igual que la relación independiente ente X e Y (Chacón (2015)).

$$convicción = \frac{n - Sop(Y)}{n(1 - conf(X \Rightarrow Y))} \quad (6)$$

3. Laplace: Para un soporte y confianza monótonas.

$$Laplace(X) = \frac{Sop(X) + 1}{Sop(X) + k} \quad (7)$$

donde la constante k es un entero mayor que 1.

4. Ganancia:

$$Gan(X) = Sop(X \Rightarrow Y) - \theta Sop(X) \quad (8)$$

5. Métrica de Piatetsky-Shapiro (P-S): Su objetivo corresponde al mismo de los anteriores.

$$P - S(X \Rightarrow Y) = Sop(X \Rightarrow Y) - \frac{Sop(X)Sop(Y)}{n} \quad (9)$$

3. Resultados

En este apartado se explican los procesamiento realizados a la base de datos y los resultados obtenidos de las reglas calculadas por el algoritmo a-priori.

3.1. Pre-procesamiento

La base de datos consta de atributos con un valor equivalente a $?$, que significa dato perdido, por lo que se reemplazaron por *NA*, también se reemplazaron los valores f y t por 0 y 1 respectivamente, esto se debe a que los atributos son caracteres y no se consideran como booleanos, mientras que con el reemplazo si se consideran.

Por consiguiente, se transforman las variables binarias a factor, para que se consideren categóricas con 2 niveles. Las variables numéricas se organizan según los rangos descritos en el laboratorio anterior, tales como, bajo, medio y alto, luego se les aplica factor, resultando en variables categóricas con tres niveles. Una excepción a las variables numéricas es la edad, pues esta se organiza en menor, medio y mayor.

Finalmente, eliminan todas las variables que contienen *measured*, esto se debe a que sus valores no afectan en los patrones, por ejemplo, si *TT4 measured* posee un valor equivalente a 0 la variable *TT4* que le corresponde sera *NA*, mientras que si *TT4 measured* tienen como valor un 1, *TT4* tendrá un valor mayor a 0 y distinto de *NA*, al igual que la variable *TBG*, se encuentra compuesta solo por valores *NA*, siendo todas las variables mencionadas consideradas valores redundantes.

3.2. Soporte de las Variables

El cuadro 1 contiene los 10 soportes más altos asociados a las variables de la base de datos

Variable	Soporte
hypopituitary = 0	0.9996
lithium = 0	0.9950
goitre = 0	0.9910
on antithyroid medication=0	0.9878
thyroid surgery=0	0.9860
pregnant=0	0.9853
I131 treatment=0	0.9828
tumor=0	0.9746
sick=0	0.9607
psych=0	0.9517

Cuadro 1: Soportes principales.

Si bien, los soportes de estas variables son elevados, no contemplan aquellas que si juegan un rol principal al momento de analizar los datos, como lo son las clases o los rangos de las variables que representan los niveles hormonales. Los cuadros 2 y 3 contienen los soportes para las clases y algunos niveles hormonales en su rango normal respectivamente.

Clase	Soporte
Negative	0.9214
Compensated	0.0550
Primary	0.0228
Secondary	0.0007

Cuadro 2: Soportes para las clases.

Hormonas	Soporte
T4U=normal	0.7839
T3=normal	0.6167
TSH=normal	0.6089
TT4=normal	0.4653
FTI=normal	0.04532

Cuadro 3: Soportes para las hormonas en su rango normal.

3.3. Soporte de la Reglas

El cuadro 4 contiene los 10 soportes más altos asociados a todas las reglas encontradas.

Antecedente	Consecuente	Soporte
on antithyroid medication=0, lithium=0, goitre=0	hypopituitary=0	0.9735
on antithyroid medication=0, goitre=0, hypopituitary=0	lithium=0	0.9735
on antithyroid medication=0, lithium=0, hypopituitary=0	goitre=0	0.9735
lithium=0, goitre=0, hypopituitary=0	on antithyroid medication=0	0.9735
thyroid surgery=0, lithium=0, goitre=0	hypopituitary=0	0.9717
thyroid surgery=0, goitre=0, hypopituitary=0	lithium=0	0.9717
thyroid surgery=0, lithium=0, hypopituitary=0	goitre=0	0.9717
lithium=0, goitre=0, hypopituitary=0	thyroid surgery=0	0.9717
pregnant=0, lithium=0, goitre=0	hypopituitary=0	0.9710
pregnant=0, goitre=0, hypopituitary=0	lithium=0	0.9710

Cuadro 4: Soportes para las reglas.

3.4. Reglas Escogidas

Un gran porcentaje de las reglas con mayor soporte terminan explicando una relación redundante sobre la información, esto debido a que las variables con mayor importancia y que permiten explicar algún fenómeno posee una incidencia extremadamente baja.

Con esto, se decidió generar un nuevo conjunto de reglas pero que cumplieran con un soporte mínimo de 0.001, puesto que el porcentaje de personas sanas corresponde a un 92 % del total de los datos, dejando solo un margen del 8 % de la información para explicar en que casos los pacientes puedan presentar alguna enfermedad. Donde incluso, ciertas clases como el hipotiriodismo secundario, representa un 0.07 % de la información.

Tanto en la figura 1 como la 2 se pueden visualizar las 100 mejores reglas según el consecuente estipulado, cabe destacar que la función implementada para graficar escoge las 100 mejores reglas, por lo que existen mas de lo expuesto en las figuras. Por otro lado, el color de los nodos indican el nivel de confianza, de modo que el rojo indica una confianza igual a 1 y a medida que decrece su valor también lo hace la saturación del color.

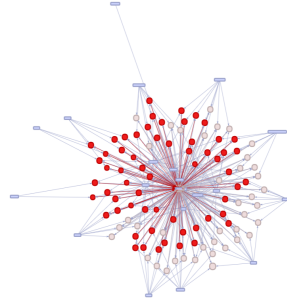


Figura 1: Reglas de asociación; consecuente = hombre

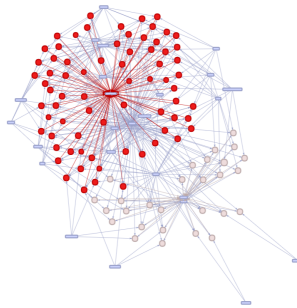


Figura 2: Reglas de asociación; consecuente = TT4 bajo

Así mismo, debido al bajo soporte existente en las reglas que pueden ser interesantes, se decide optar por aquellas que contemplen una confianza superior al 95 % y que a su vez, contengan un Lift superior a 1 para asegurar que su valor no se debe a la aleatoriedad.

El cuadro 5 contempla las reglas escogidas, donde se puede visualizar los lift's escogidos, los cuales se encuentran distantes del valor 1, a excepción de los que poseen la *clase=negative* como consecuente. Por lo tanto, podemos asegurar gracias al valor del Lift que las reglas tienden a ser reales, dado las correlaciones que existen entre los antecedentes y consecuentes.

Antecedente	Consecuente	Lift
sex=F, TSH=alto, TT4=bajo, T4U=alto	clases=primary	43.7500
TSH=alto, TT4=bajo, T4U=alto, FTI=bajo, hypopituitary=0	clases=primary	43.7500
sex=F, T3=bajo, TT4=bajo, T4U=alto	clases=primary	43.7500
sick=1, TSH=alto, T3=bajo, TT4=alto, hypopituitary=0	clases=compensated	18.1818
sick=1, I131 treatment=0, TSH=alto, T3=bajo	clases=compensated	18.1818
age=mayor, lithium=1, TT4=alto	clases=negative	1.0852
sex=F, lithium=1, TT4=alto	clases=negative	1.0852
sex=M, TSH=alto, T3=bajo, FTI=bajo	TT4=bajo	34.1463
age=mayor, sex=M, TSH=normal, T4U=bajo	TT4=bajo	2.7531
sex=F, pregnant=0, T3=bajo, clases=primary	TT4=bajo	31.1770
sex=F, on antithyroid medication=1, TT4=alto, T4U=alto	pregnant=1	68.2926
on antithyroid medication=1, tumor=0, TT4=alto, T4U=alto	pregnant=1	68.2926
on antithyroid medication=1, TT4=alto, T4U=alto, clases=negative	pregnant=1	68.2926
sick=0, goitre=0, TT4=bajo, FTI=normal	T3=bajo	12.8440

on antithyroid medication=1, pregnant=1, TT4=alto, FTI=normal	T3=alto	10.3703
I131 treatment=0, TSH=alto, TT4=bajo, cla- ses=negative	FTI=bajo	12.7853
sex=F, pregnant=0, T4U=alto, cla- ses=primary	FTI=bajo	12.7853
age=medio, pregnant=1, TSH=normal, cla- ses=negative	T4U=alto	10.6060

Cuadro 5: Reglas escogidas.

4. Análisis de Resultados

EL análisis de los resultados obtenidos es un proceso engorroso, pues según con los consecuentes estipulados la cantidad de reglas podían exceder cuatro dígitos, por ejemplo, 6039467 reglas. Por la inconveniencia mencionada, se tiene que escoger los patrones con mayor soporte y confianza, sin embargo, estos contienen variables redundantes que no aportan en la obtención de correlaciones interesantes, sino que resultan en reglas que no aportan información clara, por ejemplo, el antecedente: `on antithyroid medication=0, lithium=0, goitre=0` y consecuente: `hypopituitary=0`, no presenta como antecedente alguna característica que permita especificar si corresponde a alguna persona sana o con alguna dolencia, además de involucrar en su estudio solamente variables que podrían considerarse irrelevantes al ser características que poseen la mayoría de los individuos de una población.

De este modo, existe una gran cantidad de reglas que cumplen con un soporte elevado junto con una alta confianza, pero que su aporte es casi nulo al momento de estudiarlas, pues como se explicó en la sección anterior, los casos en los que la población si presenta las condiciones de alguna irregularidad hormonal o padecimiento, corresponden a casi un 8 % del total de las observaciones, causando que estos casos tengan un soporte extremadamente bajo, siendo el principal motivo por el cual se opto por utilizar como soporte mínimo 0.001.

4.1. Consecuente según clase

Una de las características principales de la base de datos corresponde a la clasificación que se les da a las personas que presentan o no irregularidades en sus hormonas, distinguiendo un total de cuatro clases. El Cuadro 2 indica los soportes existentes para estas clases, sin embargo, la clase Secondary no fue considerada al momento de estudiar las reglas, como se presentó en el Cuadro 5, puesto que su incidencia en la base de datos es tan baja, que sería necesario un soporte mínimo ínfimo para realizar su estudio. Los resultados del Cuadro 5 para las clases encontradas fueron los siguientes:

1. Primary: Las reglas que cumplen con esta clase como consecuente en general son aquellas en las que en su antecedente predomina un índice de TT4 bajo y TSH alto, según describe la NACB Spencer (2003). Esto ocurre porque se esperaría que las personas con

estas dos características presentaran algún tipo de hipotiroidismo al tener bajos niveles de producción de sus hormonas tiroideas.

2. Negative: En general, los individuos de esta clase corresponden a aquellos que presentan índices hormonales normales, sin embargo, resultó de interés descubrir que aquellas personas que siguen el tratamiento de litio y tengan los niveles de TT4 elevados sean considerados negativos. Existe la posibilidad de que aquellas personas que cumplan con esta características en realidad se encuentren relacionadas con el hipertiroidismo, pues el tratamiento de litio es utilizado para disminuir los niveles de T4 en quienes tienen este padecimiento Bocchetta and Loviselli (2006).
3. Compensated: Al igual que con la clase Primary su incidencia se debe al aumento de la hormona TSH, sin embargo, los niveles de T4 y T3 pueden permanecer normales o bajos. con esto, es de interés notar que exista una regla que al detectar si la persona presenta alguna enfermedad, pueda clasificarla con hipotiroidismo compensado.

4.2. Consecuente según hormona

Utilizando como guía los valores normales para las hormonas tiroideas descritos por la NACB Spencer (2003), se han tomado los casos en los que estas hormonas pueden causar un tipo de hipotiroidismo o se encuentren al limite de este, realizando el siguiente análisis:

1. TT4 bajo: Se esperaría que los individuos con esta hormona baja presenten un nivel alto de TSH y por consiguiente, una disminución en la producción de sus hormonas tiroideas. Esto se cumple en las reglas que tienen este valor como consecuente.
2. T3 bajo: El T3 se relaciona con la afectación de casi todos los procesos fisiológicos en el cuerpo, incluyendo crecimiento y desarrollo, metabolismo, temperatura corporal y ritmo cardíaco (Shahid et al. (2019a)), mientras que el TT4 consiste en la concentración de la tiroxina en el paciente (Shahid et al. (2019b)), el FTI se relaciona con los niveles libres de la tiroxina en el sistema (Shahid et al. (2019b)) y el goitre se asemeja a un crecimiento anormal de la glándula (R.A.E (2022)). Dada la regla que tiene al $T3=bajo$

como consecuente se puede inferir que el paciente no debe estar enfermo, sin crecimiento anormal de la glándula y a pesar de tener niveles de tiroxina bajo que desencadenan un nivel de T3 bajo.

3. FTI bajo: El FTI se relaciona con los niveles libres de la tiroxina en el sistema, por lo que un nivel bajo de este corresponde a un nivel bajo de tiroxina en el sistema (Shahid et al. (2019b)). Por lo tanto, una regla permite inferir que si el paciente no se encuentra en tratamiento de I131, no esta embarazada, TT4 bajo y no padece de hipotiroidismo se implica que la persona tiene un FTI bajo. Por otro lado, la segunda regla requiere de sexo femenino sin estar embarazada, un TSH normal y sin hipotiroidismo se desencadena en un nivel bajo de FTI. Ambas reglas requieren de un paciente sano, sin tener diagnosticado hipotiroidismo, pudiendo inferir que para obtener un FTI bajo, se requiere de personas sin esta enfermedad.
4. T4U alto: El T4U impide el ingreso de la tiroxina en los tejidos del cuerpo, al tener la capacidad total de unión de la hormona tiroxina a las proteína (Shahid et al. (2019b)). La única regla que posee La T4U bajo como consecuente requiere de una persona de edad media, embarazada, un TSH normal y no padecer de hipotiroidismo. Por lo tanto, se requiere de una mujer que se encuentre embarazada y sin hipotiroidismo para poder obtener un T4U alto.

4.3. Consecuente embarazadas

Las mujeres embarazadas corresponden a una población no menor de individuos que pueden verse afectados por los desordenes en sus hormonas tiroideas, llegando a desarrollar hipotiroidismo o hipertiroidismo, donde este último presenta una mayor frecuencia de aparición en ellas al aumentar sus niveles de TT4 durante la gestación. Spencer (2003) De las reglas que contienen como consecuente a las mujeres embarazadas, se observa que su antecedente contempla el rango de TT4 alto, y el uso de medicamentos antitiroideos para contrarrestar su efecto Hengameh Abdi (2019). Mas esta no es la única hormona que se encuentra en juego, pues la regla con consecuente de T3 alto, permite asegurar que al igual que el T4, las mujeres embarazadas presentan niveles elevados de esta hormona al estar de-

rechamente relacionada con los procesos fisiológicos del cuerpo. Con esto, se afirma que la tendencia de las mujeres embarazadas es desarrollar hipertiroidismo, a pesar de ser consideradas como un resultado negativo dentro de las observaciones, pues como la base de datos no contempla esta enfermedad, se le asocia el valor mas cercano que esta podría tener.

4.4. Comparación con agrupamientos

Al igual que en la experiencia anterior, la clasificación de los individuos se debe en su mayoría a las clases y los niveles hormonales, mas que a las características específicas con las que pueda verse afectada una persona. Con esto, lo esencial fue diagnosticar mediante las reglas, cuales deberían ser los antecedentes para que una persona pertenezca a una clase determinada o tenga cierto nivel hormonal, similar a lo ocurrido en la experiencia anterior, donde los individuos se agrupaban hasta tener una serie de características en común.

Con esto, los resultados obtenidos en ambas experiencias se asemejan a los descritos en la literatura, presentando un patrón similar de niveles hormonales para cada clase. en donde los niveles elevados de TSH o bajos de T4 y sus derivados indicaban la presencia o el posible padecimiento de hipotiroidismo, mientras que su contra-parte podía indicar normalidad o la presencia de hipertiroidismo, siendo el conjunto de las mujeres embarazadas quienes presentan una mayor tendencia hacia esta enfermedad.

Si bien el uso de reglas permite listar cuales son las características que un determinado grupo debería cumplir para pertenecer a una clasificación, visualmente resulta menos comprensible entender como estas características se relacionan, a comparación del método de agrupamiento, donde se distinguían claramente cuales serían los resultados obtenidos y como estos se comportaban además de tener una mayor certeza de que estos se encontraban correctos.

5. Conclusión

A partir de la experiencia fue posible alcanzar los objetivos planteados, pues se pudo generar un conjunto de reglas que permiten realizar el estudio de las variables que contempla la base de datos, para lograr obtener conocimiento a partir de ellas, al igual que la implementación del paquete estipulado.

Si bien, este objetivo se pudo completar, se esperaba que estos resultados pudieran considerar una mayor fiabilidad, pues se considera que se tuvo que disminuir demasiado el soporte aplicado a las reglas, para poder encontrar una relación que permitiese comprender su comportamiento, causando que existiera una gran cantidad de reglas que no aportaba realmente información o que contemplaran parámetros similares cayendo en la redundancia, de donde se desprende que el algoritmo utilizado es altamente sensible respecto a los datos que se le entregaban por entrada, pues una ligera variación en ellos podía generar un conjunto completamente nuevo de información.

Ahora bien, el estudio que se realizó a partir de los resultados, están acorde a lo estipulado por la literatura, además de ser muy similar a lo obtenido en la experiencia anterior dedicada al agrupamiento. Sin embargo, al comparar con la fiabilidad que se tenía en esta ultima experiencia, se retoma la misma idea del soporte implementado en las reglas, al ser tan bajo, no hay una total certeza de que todas las reglas encontradas con soporte y confianza mínima describan realmente el comportamiento del conjunto, a diferencia de lo obtenido en la experiencia anterior, donde incluso aquellos valores atípicos presentaban una asociación directa con un conjunto, haciendo que su análisis resultara mucho más sencillo.

A pesar de esto, la implementación de este método requiere de un menor pre-procesamiento de la información y visibiliza la asociación de características que mediante el agrupamiento no son posibles de detectar, gracias al uso de los antecedentes y las relaciones que estos establecen.

Para futuras experiencias, se espera implementar otros métodos que permitan comprender de manera similar la base de datos y que requieran de mayor precisión en sus resultados, pudiendo ser aseverados en los laboratorios anteriores.

Bibliografía

- Amaris, M. E. D. M., Rodríguez, J. E. R., et al. (2003). La contribución de las reglas de asociación a la minería de datos. *Tecnura*, 7(13):94–109.
- Bocchetta, A. and Loviselli, A. (2006). Lithium treatment and thyroid abnormalities. *Clinical Practice and Epidemiology in Mental Health*, 2(1):1–5.
- Chacón, D. M. (2015). Minería de datos capítulo vi “reglas de asociación”.
- Hengameh Abdi, Atieh Amouzegar, F. A. (2019). Antithyroid drugs. 18(1):1–12.
- Kotsiantis, S. and Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering*, 32(1):71–82.
- Lent, B., Swami, A., and Widom, J. (NA). Clustering association rules.
- R.A.E (2022). Bocio. <https://dle.rae.es/bocio?m=form>.
- Shahid, A., Singh, M., Raj, R., Suman, R., Jawaid, D., and Alam, M. (2019a). A study on label tsh, t3, t4u, tt4, fti in hyperthyroidism and hypothyroidism using machine learning techniques. pages 930–933.
- Shahid, A., Singh, M., Raj, R., Suman, R., Jawaid, D., and Alam, M. (2019b). A study on label tsh, t3, t4u, tt4, fti in hyperthyroidism and hypothyroidism using machine learning techniques. pages 930–933.
- Spencer, C. A. (2003). Thyroid testing for the new millenium. *Thyroid*, 13(1):2–2.
- Srikant, R. and Agrawal, R. (1995). Mining generalized association rules. pages 407–419.
- Toivonen, H., Klemettinen, M., Ronkainen, P., Hätonen, K., and Mannila, H. (1995). Pruning and grouping discovered association rules.