



## Laboratorio 3 - Reglas de Asociación

Integrantes: Christian Méndez Acosta  
Israel Arias Panez  
Curso: Análisis de Datos  
Sección A-1  
Profesor: Max Chacón Pacheco  
Ayudante: Gustavo Hurtado A.

20 de Octubre de 2022

# Tabla de contenidos

<b>1. Introducción</b>	<b>1</b>
<b>2. Marco Teórico</b>	<b>2</b>
2.1. Reglas de asociación . . . . .	2
2.2. Medidas de calidad y confianza . . . . .	2
2.2.1. Soporte . . . . .	2
2.2.2. Confianza . . . . .	2
2.2.3. Lift . . . . .	3
2.2.4. Convicción . . . . .	3
2.2.5. Cobertura . . . . .	4
2.3. Propiedades de las medidas . . . . .	4
2.4. Monotonicidad . . . . .	5
2.5. Algoritmo A priori . . . . .	5
<b>3. Pre-procesamiento</b>	<b>6</b>
<b>4. Obtención de reglas</b>	<b>9</b>
4.0.1. Reglas para clase positiva . . . . .	9
4.0.2. Reglas para clase negativa . . . . .	10
4.0.3. Reglas para variable TSH . . . . .	10
4.1. Reglas para variable T3 . . . . .	11
4.2. Reglas para variable TT4 . . . . .	11
4.3. Reglas para variable T4U . . . . .	11
4.4. Reglas para variable FTI . . . . .	12
4.5. Reglas para variable Pregnant . . . . .	12
<b>5. Análisis de resultados y comparación</b>	<b>13</b>
5.1. Reglas Clase positiva . . . . .	13
5.2. Reglas Clase negativa . . . . .	13
5.3. Reglas TSH . . . . .	14
5.4. Reglas T3 . . . . .	14
5.5. Reglas TT4 . . . . .	15

5.6. Reglas T4U . . . . .	15
5.7. Reglas FTI . . . . .	15
5.8. Reglas Pregnant . . . . .	16
5.9. Comparación con clustering . . . . .	16
<b>6. Conclusiones</b>	<b>20</b>
<b>Bibliografía</b>	<b>22</b>

# 1. Introducción

En la presente experiencia de laboratorio se busca continuar con el estudio del hipotiroidismo y la base de datos *Allhypo*. Es de importancia estudiar el hipotiroidismo, debido a como sostiene (Brent, 2022) Las hormonas tiroideas son críticas en el desarrollo somático y cerebral en infantes y regulan la actividad metabólica en adultos, también afectan a la función de todos los órganos del cuerpo.

Los objetivos para esta experiencia de laboratorio son:

- **Extraer conocimiento del hipotiroidismo, por medio de las reglas de asociación a través del software R:** Los algoritmos de reglas de asociación tienen como objetivo encontrar relaciones dentro un conjunto de transacciones, en concreto, items o atributos que tienden a ocurrir de forma conjunta (Amat, 2018). Bajo esto se busca encontrar las características que aparecen en forma conjunta, estudiarlas, determinar cuáles de estas determinan la presencia de la patología y que fenómenos se pueden identificar desde la base de datos.
- **Utilizar el paquete *arulesViz*:** El cual permitirá la obtención de reglas de asociación.
- **Contrastar la información encontrada:** La información encontrada mediante las reglas de asociación será contrastada a la información adquirida en experiencias de laboratorio anteriores, con el objetivo de corroborar o ajustar el conocimiento que se ha logrado extraer de la enfermedad.

El presente informe consta de 6 secciones, correspondientes a: introducción, marco teórico, pre-procesamiento, obtención de reglas, análisis de resultados/comparación y conclusiones. En la sección de marco teórico se presenta una breve descripción de los conceptos más relevantes para la realización y entendimiento de la experiencia. En la sección de pre-procesamiento se explicará el procesamiento realizado a los datos para poder ser trabajados bajo reglas de asociación, posteriormente en la sección de obtención de reglas se presentan las reglas obtenidas bajo distintos parámetros. En la sección de análisis de resultados y comparación se realizará el estudio de las reglas obtenidas bajo el contexto de la enfermedad, además se efectuará la comparación con los resultados de experiencias previas. Finalmente, en la sección de conclusión se realizará un resumen de lo realizado, haciendo énfasis en los principales resultados obtenidos y la comparación realizada.

## 2. Marco Teórico

### 2.1. Reglas de asociación

Una regla de asociación se define como una implicación del tipo “si X entonces Y” ( $X \Rightarrow Y$ ), donde X e Y son ítems individuales. El lado izquierdo de la regla recibe el nombre de antecedente o left-hand-side (LHS) y el lado derecho el nombre de consecuente o right-hand-side (RHS). Por ejemplo, la regla  $\{A, B\} \Rightarrow \{C\}$  significa que, cuando ocurren A y B, también ocurre C (Amat, 2018). La minería de reglas de asociación es un procedimiento cuyo objetivo es observar y encontrar patrones, correlaciones o asociaciones que se producen con frecuencia a partir de conjuntos de datos (Rai, 2022).

### 2.2. Medidas de calidad y confianza

#### 2.2.1. Soporte

El soporte de un ítem (A) es el cociente entre la cantidad de transacciones que contienen el ítem y el número total de transacciones (López et al., 2017). Este cálculo se aprecia en la Ecuación 1.

$$Soporte(A) = \frac{\text{Numero de transacciones que contienen el ítem A}}{\text{Numero de transacciones de la base de datos}} \quad (1)$$

Los autores también mencionan que el soporte de una regla representando por  $(A \rightarrow B)$  se determina por medio del cociente entre la cantidad de transacciones que contienen tanto el ítem A como el ítem B y el número total de transacciones. Esta cálculo se aprecia en la Ecuación 2.

$$Soporte(A \Rightarrow B) = \frac{\text{Numero de transacciones que contienen el ítem A y el ítem B}}{\text{Numero de transacciones de la base de datos}} \quad (2)$$

#### 2.2.2. Confianza

La confianza se interpreta como la probabilidad  $P(Y|X)$ , es decir, la probabilidad de que una transacción que contiene los ítems de X, también contenga los ítems de Y (Rai, 2022). Su definición formal se puede observar en la ecuación 3.

$$confianza(X \Rightarrow Y) = \frac{soporte(X \cup Y)}{soporte(X)} \quad (3)$$

### 2.2.3. Lift

El lift, es una medida utilizada para evaluar el grado de dependencia de los términos de una regla. En una regla de asociación  $A \Rightarrow B$ , el lift representa en qué grado “B” tiende a ser frecuente cuando “A” ocurre, o viceversa (Lucas, 2010). La medida puede ser obtenida a través de la ecuación 4.

$$lift(A \Rightarrow B) = \frac{confianza(A \Rightarrow B)}{soporte(B)} \quad (4)$$

Además, (Lucas, 2010) menciona que la evaluación de una regla de asociación puede ser hecha como se indica a continuación:

- Si  $lift(A \Rightarrow B) = 1$ , entonces la ocurrencia de los items de “B” independe de la ocurrencia de los items de “A”, y viceversa.
- Si  $lift(A \Rightarrow B) > 1$ , entonces la ocurrencia de los items de “B” influye en la probabilidad de la ocurrencia de los items de “A”.
- Si  $lift(A \Rightarrow B) < 1$ , entonces la ocurrencia de los items de “B” influye en la probabilidad de la no ocurrencia de los items de “A”.

En general, se establece que las reglas que presentan el valor de lift menor que 1 sean descartadas, como es el caso de la regla utilizada como ejemplo. Cabe destacar que (Rai, 2022) sostiene que cuanto más se aleje el valor de lift de 1, más evidencias de que la regla no se debe a un artefacto aleatorio, es decir, mayor la evidencia de que la regla representa un patrón real.

### 2.2.4. Convicción

La convicción, como sostiene (Lucas, 2010), es una medida que evalúa el grado en que el término antecedente influye en la ocurrencia del término consecuente de una regla de asociación. A diferencia del lift, convicción es una medida unidireccional, o sea, el resultado de  $conviccion(A \Rightarrow B)$  va a ser diferente del de  $conviccion(B \Rightarrow A)$ . La Convicción puede ser obtenida como se describe en la ecuación 5.

$$conviccion(A \Rightarrow B) = \frac{soporte(A, \neg B)}{soporte(A) \times soporte(\neg B)} \quad (5)$$

El valor de la convicción puede variar entre 1 y  $+\infty$ . Cuanto más alto el valor, mayor va a ser la probabilidad de ocurrencia del término consecuente cuando el término antecedente de la regla ocurra. El valor 1 indica la independencia de los términos de la regla.

### 2.2.5. Cobertura

Respecto a la medida de cobertura o coverage (Lucas, 2010), menciona que expresa la proporción de objetos o registros en determinado conjunto de datos que son cubiertos por los ítems del término antecedente de una determinada regla. De esta forma, se puede definir variadas maneras de obtener dicha medida. La cobertura puede ser obtenida a través del propio soporte de una regla, como se puede ver en la ecuación 6.

$$cobertura(A \Rightarrow B) = soporte(A) \quad (6)$$

## 2.3. Propiedades de las medidas

Ciertos conjuntos de reglas pueden cumplir ciertas características, las cuales serán descritas a continuación:

- Reglas confiables: Se entiende como el conjunto de todas las reglas que cumplan con una confianza mínima (Chacón, 2015).
- Reglas frecuentes: Se entiende como el conjunto de todas las reglas que cumplan con un soporte mínimo (Chacón, 2015).
- Reglas interesantes: Se entiende como el conjunto de todas las reglas que son confiables y frecuentes bajo una confianza y soporte mínimo dados (Chacón, comunicación personal, octubre de 2022).
- Especialización: Si la búsqueda de los posibles A conjuntos se realiza agregando condiciones se denomina especialización de la regla (Chacón, 2015).
- Generalización: Si el antecedente de la regla A1, contiene menos condiciones que la regla A2, se dice que A1 es más general que A2 (Chacón, 2015).

## 2.4. Monotonicidad

El carácter monótono involucra que cierto comportamiento esta perdurando, esta propiedad restringe la búsqueda de reglas que cumplan ciertas retriicciones, en función de la especialización. Se dice que la medida o restricción es monótona si:  $med(A1) \leq med(A2)$ . La medida es anti-monótona si:  $med(A1) \geq med(A2)$ . Para ejemplificar esto, si se utiliza el soporte como medida se puede observar que es anti-monótona, ya que la especialización de la regla lleva a mantener o disminuir el soporte. Por otra parte, si se utiliza la confianza como medida, al especializar la regla se puede mantener o disminuir el soporte, además el soporte del antecedente puede disminuir en una proporción mayor que el soporte de la regla, al ser la confianza la razón entre estos, provocaría un aumento en la confianza (Chacón, 2015).

## 2.5. Algoritmo A priori

A priori fue uno de los primeros algoritmos desarrollados para la búsqueda de reglas de asociación, el algoritmo realiza una búsqueda exhaustiva por niveles de complejidad (de menor a mayor tamaño de itemsets), para limitar la explosión combinatoria de posibilidades se aprovecha del principio de monotonidad, como tal el algoritmo consta de dos etapas:

1. Identificar todos los itemsets que ocurren con una frecuencia por encima de un determinado límite (itemsets frecuentes).
2. Convertir esos itemsets frecuentes en reglas de asociación.

Dado que se tiene un carácter monótono, si un conjunto es infrecuente, entonces, todos los conjuntos donde este último se encuentre, también son infrecuentes, bajo esto se realiza una poda (Amat, 2018).



### 3. Pre-procesamiento

Al igual que para la experiencia de laboratorio anterior, se hace necesario el efectuar un pre-procesamiento de los datos antes de poder empezar a trabajar con el método a aplicar. Al igual que en la experiencia previa, se trabaja con la base de datos *Allhypo* la cual fue donada dentro de un grupo de distintas bases de datos con registros sobre enfermedades a la tiroides aportados por el instituto de Garavan y la Universidad de California (Dua and Graff, 2017). Esta base de datos contiene 2800 observaciones y 30 variables, sin embargo, al realizar una exploración inicial de los datos es posible observar que existen variables que no aportan al estudio por mencionar si cierta hormona fue medida o no, estas son: TSH measured, T3 measured, TT4 measured, T4U measured, FTI measured, TBG measured. Sumadas a estas la variable referral source hace alusión al origen de donde proviene la observación, lo que tampoco aporta al estudio. Así mismo la variable hypopituitary no presenta varianza, al tener solo una muestra con un valor distinto a false, por lo que no aporta al estudio y por último la variable TBG no fue registrada en ninguna de las observaciones, por lo que tampoco aporta información. Estas 9 variables son eliminadas de la base de datos, posterior a esto se procedió a observar los datos faltantes que presenta la base de datos, en la Figura 1 es posible visualizar el resumen de los datos faltantes, donde es posible apreciar que se presentan datos faltantes en las variables T3, T4U, FTI, TSH, TT4, sex y age. Respecto a datos faltantes (Lodder et al., 2013) sostienen que la regla general es que si falta menos del 5-10% de los datos, la eliminación de muestras ya no supone una amenaza importante para la potencia estadística. Por otro lado, (Alice, 2015) recomienda que, si los datos que faltan para una determinada muestra no son superiores al 5%, debería dejar fuera esa muestra. Tomando en consideración estos factores, se decidió el eliminar las muestras que presentan menos del 5% de datos faltantes, como ocurre en las variables sex y age.

Para tratar los datos faltantes con perdida mayor al 5% se tomó la decisión de imputar los datos faltantes, esto ocurre en las variables T3, T4U, FTI, TSH y TT4. Para la imputación de los datos se decidió usar el método de predictive mean matching, como menciona (Allison, 2015), usar predictive mean matching es una forma atractiva de realizar la imputación de múltiples datos faltantes. En comparación con los métodos estándar basados en la regresión lineal y la distribución normal, predictive mean matching produce valores imputados que se parecen mucho más a los valores reales. El método de predictive

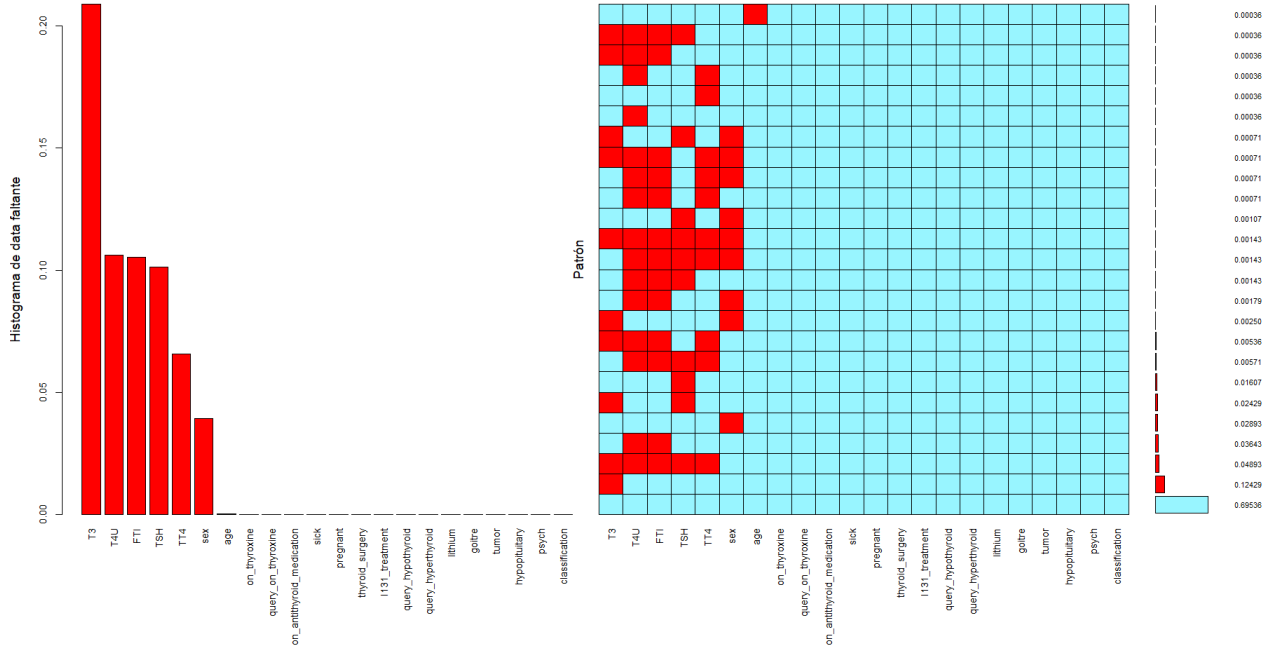


Figura 1: Histograma y gráfica de patrón de datos faltantes

mean matching funciona haciendo coincidir la distancia media predictiva de las observaciones incompletas con las de las observaciones completas (Akmam et al., 2019).

Luego de estos procedimientos efectuados, la base de datos presenta 21 variables y 2688 observaciones útiles para el estudio del problema.

Para poder trabajar con reglas de asociación la base de datos no puede contener variables numéricas, como menciona (Amat, 2018), las técnicas de reglas de asociación solo trabajan con información booleana, es decir, con la presencia o no de cada uno de los items en la transacción. Por lo tanto, si el set de datos contiene variables numéricas, estas deben de ser discretizadas en intervalos o niveles. Ante esto, las variables numéricas: TSH, T3, TT4, T4U y FTI fueron discretizadas en: Baja, Normal y Alta de acuerdo a sus rangos de referencia. Los rangos de referencia para las hormonas T3, TSH y TT4 fueron obtenidos según lo sostenido por (Gonzalez, 2021). En el caso de las hormonas T4U y FTI los rangos de referencia ocupados fueron los presentados por (Symons and Vining, 1985). El resumen de la discretización aplicada puede ser observado en la Tabla 1.

En el caso de la edad, está fue discretizada tomando en consideración una versión resumida de las etapas de la vida presentadas por (Ministerio de salud y protección social, sf). El resumen de esta discretización puede observarse en la Tabla 2.

Variable	Rango		
	Baja	Normal	Alta
TSH	$-\infty - 0.36$	0.37 - 4.7	4.8 - $+\infty$
T3	$-\infty - 1.1$	1.2 - 2.7	2.8 - $+\infty$
TT4	$-\infty - 59$	60 - 150	151 - $+\infty$
T4U	$-\infty - 0.73$	0.74 - 1.08	1.09 - $+\infty$
FTI	$-\infty - 87$	88 - 132	133 - $+\infty$

Tabla 1: Discretización de variables relacionadas a hormonas

Rango	
Infancia	0 - 11
Juventud	12- 26
Adultez	27 - 59
Vejez	60 - $+\infty$

Tabla 2: Discretización de variable edad

Por último, se analizó la cantidad de observaciones de cada clase disponibles en la base de datos, la cual corresponde a: 92.15 % de la clase negative, 5.54 % de la clase compensated hypothyroid, 2.23 % de la clase primary hypothyroid y 0.07 % de la clase secondary hypothyroid. Es posible darse cuenta de que las diferencias entre la cantidad de observaciones de la clase negativa y las clases que presentan un tipo de hipotiroidismo es muy notable, es debido a esto y con el fin de aumentar el soporte de las reglas a obtener para la presencia de la enfermedad, que las clases relacionadas a la presencia de hipotiroidismo (primary hypothyroid, compensated hypothyroid y secondary hypothyroid) son unidas en una sola clase llamada positive.

## 4. Obtención de reglas

Con el fin de obtener reglas que aporten información al problema, estas se obtendrán usando como consecuente a las clases, además de las variables más relevantes observadas en la primera experiencia de laboratorio (Arias and Méndez, 2022).

Una de las principales dificultades al aplicar minería de reglas de asociación es el fijar un valor de soporte mínimo. Un alto valor evita la explosión combinatoria, pero a la expensa de no descubrir reglas o patrones interesantes que tienen bajo soporte. Sin embargo, muchas reglas con alto soporte son obvias o bien conocidas y es en las reglas con bajo soporte en donde se consiguen nuevas e interesantes perspectivas como desviaciones o excepciones. Todos los enfoques actuales ponen al usuario en un dilema: ¿Cómo elegir el valor apropiado de soporte mínimo para descubrir información relevante? Ante esto lo mejor que se puede hacer es fijar el valor de soporte al mínimo posible, de forma de no perder ningún caso relevante y previo a esto efectuar un filtro que permita dejar todas las reglas con lift mayor a 1 y con la confianza escogida. Con esto se asegura conseguir todas las reglas interesantes, este procedimiento se llama especificación de confidence-lift (Lin and Tseng, 2006). Se ha decidido seguir este enfoque, ya que, a pesar de generar muchas combinaciones, es el recomendado en la literatura y el ocupado en otras investigaciones, además (Selvi and Tamilarasi, 2009) mencionan en su investigación que los resultados experimentales mostraron que el modelo propuesto basado en la especificación de confidence-lift rinde bien y genera reglas relevantes sin perder reglas interesantes. La confianza mínima para la obtención de reglas a usar será la mayor que se pueda obtener para cada consecuente como recomienda la especificación de confidence-lift, además a esto se ha limitado la generación de reglas a que incluyan al menos dos antecedentes, debido a que las reglas con un solo ítem de antecedente no aportan mucha información cuando el interés es estudiar el comportamiento de una enfermedad.

### 4.0.1. Reglas para clase positiva

A continuación se presentan las reglas para la clase positiva (presencia de la enfermedad), de acuerdo a los parámetros determinados al inicio de esta sección. Las mejores 5 reglas según confianza y lift se pueden ver en la Figura 2.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{thyroid_surgery=f, TSH=Alta, T3=Baja, TT4=Baja}	=> {classification=positive}	0.01227679	0.9428571	0.01302083	12.01137	33
[2]	{thyroid_surgery=f, TSH=Alta, T3=Baja, TT4=Baja, FTI=Baja}	=> {classification=positive}	0.01227679	0.9428571	0.01302083	12.01137	33
[3]	{thyroid_surgery=f, psych=f, TSH=Alta, T3=Baja, TT4=Baja}	=> {classification=positive}	0.01227679	0.9428571	0.01302083	12.01137	33
[4]	{sick=f, thyroid_surgery=f, TSH=Alta, T3=Baja, TT4=Baja}	=> {classification=positive}	0.01227679	0.9428571	0.01302083	12.01137	33
[5]	{thyroid_surgery=f, tumor=f, TSH=Alta, T3=Baja, TT4=Baja}	=> {classification=positive}	0.01227679	0.9428571	0.01302083	12.01137	33

Figura 2: Mejores reglas encontradas para clase positiva

#### 4.0.2. Reglas para clase negativa

A continuación se presentan las reglas obtenidas para la clase negativa (no posee enfermedad tiroidea), los parámetros utilizados son los mismos que en la clase positiva. Las mejores 5 reglas según confianza y lift se pueden ver en la Figura 3

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{TSH=Normal, FTI=Normal}	=> {classification=negative}	0.4665179	1	0.4665179	1.085184	1254
[2]	{TSH=Normal, TT4=Normal}	=> {classification=negative}	0.5825893	1	0.5825893	1.085184	1566
[3]	{TSH=Normal, T4U=Normal, FTI=Normal}	=> {classification=negative}	0.3422619	1	0.3422619	1.085184	920
[4]	{TSH=Normal, T3=Normal, FTI=Normal}	=> {classification=negative}	0.3917411	1	0.3917411	1.085184	1053
[5]	{TSH=Normal, TT4=Normal, FTI=Normal}	=> {classification=negative}	0.4471726	1	0.4471726	1.085184	1202

Figura 3: Mejores reglas encontradas para clase negativa

#### 4.0.3. Reglas para variable TSH

Como se mencionó al inicio de esta sección, además de la clase también se incluye el estudio de variables que han sido considerado interesantes en el estudio de la enfermedad. Como lo son las hormonas tiroideas. En la figura 4 se incluyen las 5 reglas más interesantes encontradas para la hormona TSH cuando esta es alta.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{sex=F, on_thyroxine=f, query_hyperthyroid=f, TT4=Baja, FTI=Baja}	=> {TSH=Alta}	0.01264881	0.7555556	0.01674107	5.610313	34
[2]	{sex=F, on_thyroxine=f, query_hyperthyroid=f, psych=f, TT4=Baja, FTI=Baja}	=> {TSH=Alta}	0.01264881	0.7555556	0.01674107	5.610313	34
[3]	{sex=F, on_thyroxine=f, sick=f, query_hyperthyroid=f, TT4=Baja, FTI=Baja}	=> {TSH=Alta}	0.01264881	0.7555556	0.01674107	5.610313	34
[4]	{sex=F, on_thyroxine=f, l131_treatment=f, query_hyperthyroid=f, TT4=Baja, FTI=Baja}	=> {TSH=Alta}	0.01264881	0.7555556	0.01674107	5.610313	34
[5]	{sex=F, on_thyroxine=f, pregnant=f, query_hyperthyroid=f, TT4=Baja, FTI=Baja}	=> {TSH=Alta}	0.01264881	0.7555556	0.01674107	5.610313	34

Figura 4: Mejores reglas encontradas para TSH

Cabe destacar que tanto en este caso como en los próximos relacionados a variables no fue considerada la clase en el antecedente, debido a que no tendría sentido que la clase este prediciendo una variable.

## 4.1. Reglas para variable T3

En la figura 5 es posible observar las 5 reglas más interesantes encontradas para la hormona T3 cuando esta es baja.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{age=Vejez, on_thyroxine=f, query_on_thyroxine=f, on_antithyroid_medication=f, sick=f, goitre=f, T4U=Baja}	=> {T3=Baja}	0.01153274	0.8157895	0.0141369	8.499388	31
[2]	{age=Vejez, on_thyroxine=f, query_on_thyroxine=f, on_antithyroid_medication=f, sick=f, query_hyperthyroid=f, goitre=f, T4U=Baja}	=> {T3=Baja}	0.01153274	0.8157895	0.0141369	8.499388	31
[3]	{age=Vejez, on_thyroxine=f, query_on_thyroxine=f, on_antithyroid_medication=f, sick=f, I131_treatment=f, goitre=f, T4U=Baja}	=> {T3=Baja}	0.01153274	0.8157895	0.0141369	8.499388	31
[4]	{age=Vejez, on_thyroxine=f, query_on_thyroxine=f, on_antithyroid_medication=f, sick=f, pregnant=f, goitre=f, T4U=Baja}	=> {T3=Baja}	0.01153274	0.8157895	0.0141369	8.499388	31
[5]	{age=Vejez, on_thyroxine=f, query_on_thyroxine=f, on_antithyroid_medication=f, sick=f, thyroid_surgery=f, goitre=f, T4U=Baja}	=> {T3=Baja}	0.01153274	0.8157895	0.0141369	8.499388	31

Figura 5: Mejores reglas encontradas para T3

## 4.2. Reglas para variable TT4

En la figura 6 se presentan las mejores 5 reglas más interesantes para la hormona TT4 cuando esta es baja.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{on_thyroxine=f, query_on_thyroxine=f, I131_treatment=f, psych=f, TSH=Alta, T3=Baja, FTI=Baja}	=> {TT4=Baja}	0.01264881	0.8717949	0.01450893	26.03761	34
[2]	{on_thyroxine=f, query_on_thyroxine=f, sick=f, I131_treatment=f, TSH=Alta, T3=Baja, FTI=Baja}	=> {TT4=Baja}	0.01264881	0.8717949	0.01450893	26.03761	34
[3]	{on_thyroxine=f, query_on_thyroxine=f, sick=f, I131_treatment=f, psych=f, TSH=Alta, T3=Baja, FTI=Baja}	=> {TT4=Baja}	0.01264881	0.8717949	0.01450893	26.03761	34
[4]	{on_thyroxine=f, query_on_thyroxine=f, I131_treatment=f, tumor=f, psych=f, TSH=Alta, T3=Baja, FTI=Baja}	=> {TT4=Baja}	0.01264881	0.8717949	0.01450893	26.03761	34
[5]	{on_thyroxine=f, query_on_thyroxine=f, pregnant=f, I131_treatment=f, psych=f, TSH=Alta, T3=Baja, FTI=Baja}	=> {TT4=Baja}	0.01264881	0.8717949	0.01450893	26.03761	34

Figura 6: Mejores reglas encontradas para TT4

## 4.3. Reglas para variable T4U

En la figura 7 se presentan las mejores 5 reglas más interesantes para el índice T4U cuando este es bajo.

lhs	rhs	support	confidence	coverage	lift	count
[1] {sick=f, query_hypothyroid=f, query_hyperthyroid=f, TSH=Normal, T3=Baja}	=> {T4U=Baja}	0.01004464	0.2389381	0.04203869	5.683765	27
[2] {sick=f, I131_treatment=f, query_hypothyroid=f, query_hyperthyroid=f, TSH=Normal, T3=Baja}	=> {T4U=Baja}	0.01004464	0.2389381	0.04203869	5.683765	27
[3] {sick=f, pregnant=f, query_hypothyroid=f, query_hyperthyroid=f, TSH=Normal, T3=Baja}	=> {T4U=Baja}	0.01004464	0.2389381	0.04203869	5.683765	27
[4] {sick=f, thyroid_surgery=f, query_hypothyroid=f, query_hyperthyroid=f, TSH=Normal, T3=Baja}	=> {T4U=Baja}	0.01004464	0.2389381	0.04203869	5.683765	27
[5] {on_antithyroid_medication=f, sick=f, query_hypothyroid=f, query_hyperthyroid=f, TSH=Normal, T3=Baja}	=> {T4U=Baja}	0.01004464	0.2389381	0.04203869	5.683765	27

Figura 7: Mejores reglas encontradas para T4U

#### 4.4. Reglas para variable FTI

En la figura 8 se presentan las mejores 5 reglas más interesantes para la hormona FTI cuando esta es baja.

lhs	rhs	support	confidence	coverage	lift	count
[1] {TT4=Baja, T4U=Normal}	=> {FTI=Baja}	0.01748512	1	0.01748512	5.530864	47
[2] {TSH=Alta, T3=Baja, TT4=Baja}	=> {FTI=Baja}	0.01339286	1	0.01339286	5.530864	36
[3] {age=Adultez, TSH=Alta, TT4=Baja}	=> {FTI=Baja}	0.01041667	1	0.01041667	5.530864	28
[4] {sex=F, TSH=Alta, TT4=Baja}	=> {FTI=Baja}	0.01450893	1	0.01450893	5.530864	39
[5] {TSH=Alta, TT4=Baja, T4U=Normal}	=> {FTI=Baja}	0.01078869	1	0.01078869	5.530864	29

Figura 8: Mejores reglas encontradas para FTI

#### 4.5. Reglas para variable Pregnant

En la figura 9 se presentan las mejores 5 reglas más interesantes para la variable Pregnant cuando esta es verdadera, o sea cuando la persona se encuentra embarazada.

lhs	rhs	support	confidence	coverage	lift	count
[1] {sex=F, sick=f, I131_treatment=f, query_hypothyroid=f, T3=Alta, T4U=Alta}	=> {pregnant=t}	0.01078869	0.1972789	0.0546875	13.25714	29
[2] {sex=F, sick=f, thyroid_surgery=f, I131_treatment=f, query_hypothyroid=f, T3=Alta, T4U=Alta}	=> {pregnant=t}	0.01078869	0.1972789	0.0546875	13.25714	29
[3] {sex=F, sick=f, I131_treatment=f, query_hypothyroid=f, goitre=f, T3=Alta, T4U=Alta}	=> {pregnant=t}	0.01078869	0.1972789	0.0546875	13.25714	29
[4] {sex=F, sick=f, I131_treatment=f, query_hypothyroid=f, lithium=f, T3=Alta, T4U=Alta}	=> {pregnant=t}	0.01078869	0.1972789	0.0546875	13.25714	29
[5] {sex=F, sick=f, thyroid_surgery=f, I131_treatment=f, query_hypothyroid=f, goitre=f, T3=Alta, T4U=Alta}	=> {pregnant=t}	0.01078869	0.1972789	0.0546875	13.25714	29

Figura 9: Mejores reglas encontradas para Pregnant

Además a todas las reglas generadas, se intentó conseguir reglas usando como consecuentes para las variables goitre y lithium, dado lo descubierto sobre ellas en la primera experiencia de laboratorio (Arias and Méndez, 2022). Sin embargo no fue posible generar alguna regla debido a no cumplir con el mínimo de soporte de 0.01.

## 5. Análisis de resultados y comparación

En esta sección se llevará a cabo el análisis de resultados de las reglas conseguidas en la sección anterior, además de una comparación con los resultados de la experiencia de laboratorio anterior. Cabe destacar que todas las reglas conseguidas siguen la especificación de confidence-lift (Lin and Tseng, 2006), por lo que se asegura un lift mayor a 1 y la mayor confianza que se puede obtener para el consecuente a analizar. A continuación, se analizarán todas las reglas conseguidas por consecuente.

### 5.1. Reglas Clase positiva

En la Figura 2, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es la clase positiva, cabe destacar que en cada uno de los antecedentes se encuentran los siguiente niveles hormonales: **TSH=Alta**, **T3=Baja** y **TT4=Baja**. Estos niveles hormonales coinciden con lo estipulado en la literatura para pacientes que padecen algún tipo de hipotiroidismo, ya que el hipotiroidismo es detectado ya sea por niveles altos de TSH o niveles bajos de tiroxina (T4) (Mayo Clinic, sf). Otro antecedente que se repite en las reglas obtenidas es no contar con cirugía de tiroides (thyroid\_surgery=f), esto resulta interesante debido a que el tratamiento para pacientes con hipotiroidismo no suele ser a través de una cirugía sino por medio de un tratamiento farmacológico, en concreto se realiza una suplementación de tiroxina (T4) (Galofre, sf).

Todas las reglas poseen una confianza de alrededor del 94 %, por lo cual, se puede considerar como verdaderas las causas que provocan la clasificación de la clase positiva, esto se refuerza al observar que las reglas obtenidas se ven respaldadas por la literatura.

### 5.2. Reglas Clase negativa

En la Figura 3, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es la clase negativa, el antecedente de las reglas se compone de las variables TSH, T3, TT4, T4U y FTI, cabe destacar que únicamente se tienen niveles hormonales normales al predecir la clase negativa y la hormona TSH aparece en cada una de las reglas, lo cual tiene sentido, ya que como se mencionó anteriormente según (Mayo Clinic, sf), el hipotiroidismo (clase positiva) es detectado por niveles altos de TSH o niveles bajos de tiroxina (T4).



En este caso las reglas presentadas poseen un 100 % de confianza, por lo cual, las causas que predicen la clasificación negativa parecen ser verdaderas, esto se refuerza al observar que las reglas obtenidas se ven respaldadas por la literatura.

### 5.3. Reglas TSH

En la Figura 4, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es TSH=Alta, cada uno de los antecedentes posee sexo femenino, on\_thyroxine=f, query\_hyperthyroid=f y niveles de TT4 y FTI bajos. Los niveles bajos de T4 indican que el paciente puede tener alguna clase de hipotiroidismo, la clase de hipotiroidismo que posee el paciente es detectada por medio de los niveles de TSH que posea, en caso de tener niveles altos de TSH el paciente podría padecer de hipotiroidismo primario, por otra parte, niveles bajos de TSH podría indicar que el paciente padece de hipotiroidismo secundario (Tu otro médico, 2020).

Otro aspecto interesante de los antecedentes es la presencia del sexo femenino, ya que, al retomar la idea expuesta en la clase positiva, niveles altos de TSH y niveles bajos de T4 indicarían que los pacientes poseerían algún tipo de hipotiroidismo, condición que suele ser más frecuente en mujeres a partir de los 40 años (Galofre, sf).

En este caso las reglas presentadas poseen alrededor del 75 % de confianza, por lo cual, las causas que predicen niveles altos de TSH parecen ser adecuadas, esto se refuerza al observar que las reglas obtenidas se ven respaldadas por la literatura.

### 5.4. Reglas T3

En la Figura 5, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es T3=Baja, cada uno de los antecedentes posee age=Vejez, on\_thyroxine=f, query\_on\_thyroxine=f, on\_antithyroid\_medication=f, sick=f, goitre=f, T4U=Baja. Esta regla agrupó a los pacientes que no poseen alguna condición especial o estén en algún tratamiento para la tiroides. Las variables más interesantes son la edad (age=Vejez) y el radio de utilización de tiroxina (T4U=Baja), respecto a los niveles hormonales en adultos mayores (Montero, 2009), menciona que las alteraciones en los niveles hormonales en adultos mayores se pueden presentar de cualquier forma, sin embargo las más comunes son niveles de T3 y T4 bajos acompañados de niveles normales o altos de TSH.

La confianza de las reglas obtenidas se encuentra alrededor del 81 %, por lo cual, las causas que predicen niveles bajos de T3 parecen ser adecuadas, esto se refuerza al observar que las reglas obtenidas se ven respaldadas por la literatura.

### 5.5. Reglas TT4

En la Figura 6, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es TT4=Baja, cada uno de los antecedentes posee `on_thyroxine=f`, `query_on_thyroxine=f`, `I131_treatment=f`, `TSH=Alta`, `T3=Baja` y `FTI=Baja`. Estos pacientes no se encuentran en algún tipo de tratamiento, además como mencionó (Tu otro médico, 2020), los niveles hormonales altos en TSH y bajos en T4 y T3 indicarían pacientes con hipotiroidismo primario.

Las reglas obtenidas poseen una confianza que ronda en un 87 %, por ende, las causas que predicen niveles bajos de TT4 parecen ser acertadas, esto se refuerza al observar que las reglas obtenidas se ven respaldadas por la literatura.

### 5.6. Reglas T4U

En la Figura 7, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es T4U=Baja, cada uno de los antecedentes se compone de `sick=f`, `query_hyperthyroid=f`, `query_hypothyroid=f`, `TSH=Normal` y `T3=Baja`. Los pacientes poseen niveles normales de TSH pero bajos en T3 y utilización de T4, estos niveles son indicadores de que el paciente puede padecer hipotiroidismo secundario, ya que este se detecta por medio de disminuciones en niveles de T3, leve disminución o normalidad de los niveles de T4 y normalidad de los niveles de TSH (Villanueva, 2001).

En este caso las reglas presentadas poseen alrededor del 23 % de confianza, por lo cual, las causas que predicen niveles bajos de T4U resultan inciertas, a pesar de esta baja confianza la literatura respalda las reglas obtenidas.

### 5.7. Reglas FTI

En la Figura 8, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es FTI=Baja, en este caso la única variable que se repite en todos los antecedentes

es la cantidad total de hormona T4 en el paciente (TT4=Baja), en cuanto al resto de reglas el nivel de TSH alto indicaría que el paciente podría padecer un hipotiroidismo primario (Tu otro médico, 2020). La aparición de la edad (Adultez) y el sexo (femenino) refuerza la idea previamente expuesta de que las mujeres sobre 40 años poseen una mayor probabilidad de padecer hipotiroidismo (American Thyroid Association, sf).

Las reglas obtenidas poseen una confianza de un 100 %, por ende, las causas que predicen niveles bajos de FTI parecen ser verdaderas, esto se refuerza al observar que las reglas obtenidas se ven respaldadas por la literatura.

## 5.8. Reglas Pregnant

En la Figura 9, se pueden apreciar las 5 mejores reglas encontradas cuando el consecuente es `pregnant=t`, las pacientes no presentan alguna condición extra o aplicación de algún tratamiento, sin embargo los niveles hormonales de T3 y T4U aumentan durante el embarazo, ante esta alteración (American Thyroid Association, sf), comenta que durante el embarazo o la utilización de píldoras anticonceptivas se pueden provocar aumentos en los niveles de T4 y T3 totales.

En este caso las reglas presentadas poseen alrededor del 19 % de confianza, a pesar de esta baja confianza la literatura respalda el aumento en el nivel hormonal de T3 y T4U en mujeres embarazadas.

## 5.9. Comparación con clustering

A continuación, se realizará una comparación entre los resultados obtenidos en esta experiencia de laboratorio con las obtenidas en la experiencia de laboratorio anterior, en la cual se trabajó con clustering. En la Figura 10 es posible observar el cluster con el cual se obtuvieron los mejores resultados en la experiencia anterior (Arias and Méndez, 2022). Además es posible observar en las Figuras 11, 12 y 13 las tablas resumen que muestran lo que contiene cada cluster.

Como primera comparación y como se mencionó en la experiencia anterior (Arias and Méndez, 2022), cabe recordar que el cluster 3 (de color azul en la Figura 10) agrupa a los pacientes con hipotiroidismo primario, su tabla resumen se presenta en la Figura 13. Una persona con hipotiroidismo primario se caracteriza por tener niveles hormonales de TSH

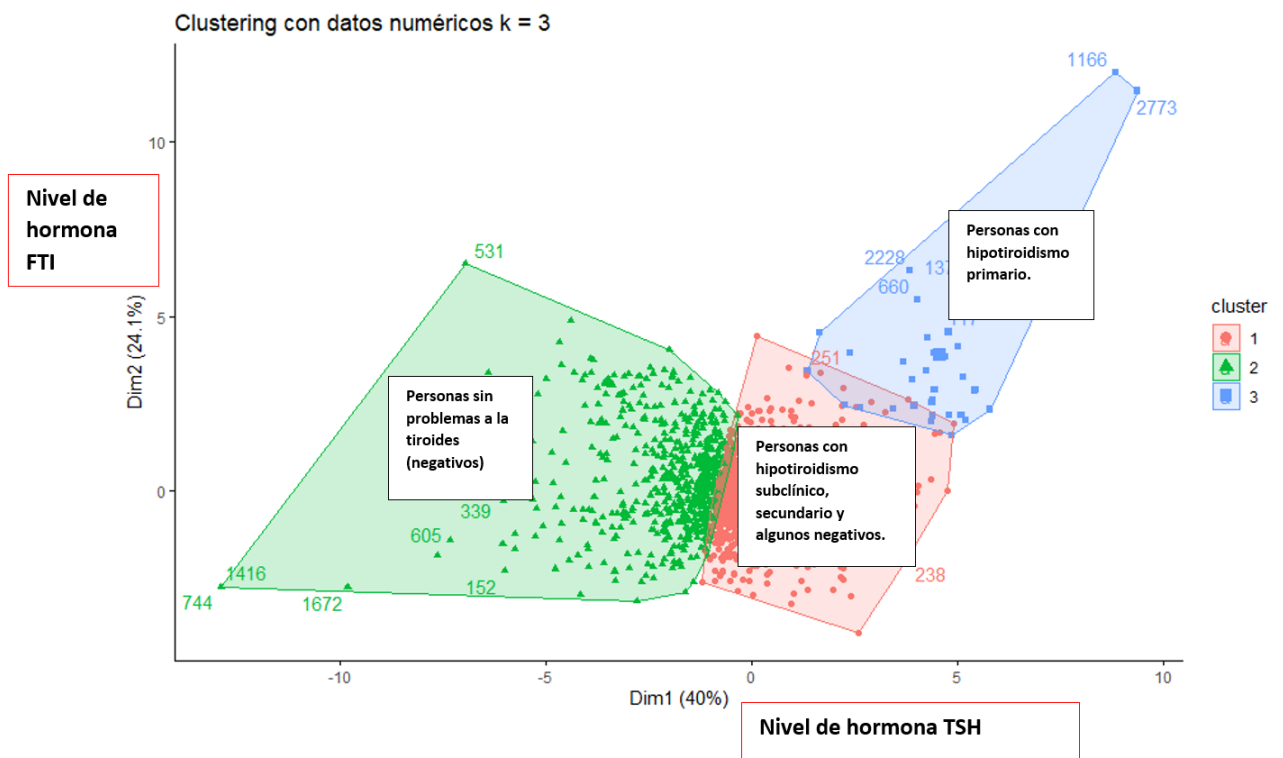


Figura 10: Cluster Experiencia Laboratorio 2, K = 3

```
[[1]]
```

age	TSH	T3	TT4	T4U	FTI	cluster
Min. : 1.00	Min. : 0.005	Min. : 0.050	Min. : 3.00	Min. : 0.3100	Min. : 3.0	Min. : 1
1st Qu.: 39.00	1st Qu.: 0.640	1st Qu.: 1.500	1st Qu.: 84.00	1st Qu.: 0.8600	1st Qu.: 90.0	1st Qu.: 1
Median : 57.00	Median : 1.500	Median : 1.900	Median : 98.00	Median : 0.9400	Median : 103.0	Median : 1
Mean : 53.82	Mean : 3.108	Mean : 1.825	Mean : 97.08	Mean : 0.9476	Mean : 103.6	Mean : 1
3rd Qu.: 69.00	3rd Qu.: 2.800	3rd Qu.: 2.200	3rd Qu.: 111.00	3rd Qu.: 1.0400	3rd Qu.: 117.0	3rd Qu.: 1
Max. : 94.00	Max. : 58.000	Max. : 4.800	Max. : 165.00	Max. : 1.6800	Max. : 189.0	Max. : 1

```
clasificacion
```

compensated hypothyroid:	136
negative	:1911
primary hypothyroid	: 32
secondary hypothyroid	: 2

Figura 11: Tabla resumen cluster N°1

```
[[2]]
```

age	TSH	T3	TT4	T4U	FTI	cluster
Min. : 2.00	Min. : 0.005	Min. : 1.100	Min. : 90.0	Min. : 0.640	Min. : 58.0	Min. : 2
1st Qu.: 29.00	1st Qu.: 0.065	1st Qu.: 2.300	1st Qu.: 134.0	1st Qu.: 0.990	1st Qu.: 109.0	1st Qu.: 2
Median : 41.50	Median : 0.540	Median : 2.600	Median : 149.0	Median : 1.130	Median : 131.0	Median : 2
Mean : 44.44	Mean : 1.338	Mean : 2.899	Mean : 156.7	Mean : 1.181	Mean : 138.5	Mean : 2
3rd Qu.: 60.00	3rd Qu.: 1.700	3rd Qu.: 3.400	3rd Qu.: 170.0	3rd Qu.: 1.310	3rd Qu.: 159.0	3rd Qu.: 2
Max. : 94.00	Max. : 26.000	Max. : 10.600	Max. : 430.0	Max. : 2.120	Max. : 395.0	Max. : 2

```
clasificacion
```

compensated hypothyroid:	12
negative	:560
primary hypothyroid	: 0
secondary hypothyroid	: 0

Figura 12: Tabla resumen cluster N°2

[[3]]							
age	TSH	T3	TT4	T4U	FTI	cluster	
Min. :14.00	Min. : 47.0	Min. :0.2000	Min. : 2.00	Min. :0.760	Min. : 2.00	Min. :3	
1st Qu.:36.50	1st Qu.: 80.0	1st Qu.:0.4000	1st Qu.: 12.50	1st Qu.:0.995	1st Qu.:11.00	1st Qu.:3	
Median :50.00	Median :103.0	Median :0.7000	Median : 27.00	Median :1.100	Median :26.00	Median :3	
Mean :48.69	Mean :132.1	Mean :0.9086	Mean : 34.67	Mean :1.093	Mean :32.28	Mean :3	
3rd Qu.:61.00	3rd Qu.:151.0	3rd Qu.:1.3500	3rd Qu.: 54.00	3rd Qu.:1.185	3rd Qu.:49.50	3rd Qu.:3	
Max. :79.00	Max. :478.0	Max. :2.1000	Max. :106.00	Max. :1.500	Max. :91.00	Max. :3	
clasificacion							
compensated hypothyroid: 1							
negative : 6							
primary hypothyroid :28							
secondary hypothyroid : 0							

Figura 13: Tabla resumen cluster N°3

alta y FTI (Tiroxina libre) baja (Gonzalez, 2021). Las medias de las hormonas presentes en este cluster son de: TSH = 132.1, T3 = 0.9086, TT4 = 34.67 y FTI = 32.28. Al comparar estos valores con los rangos de referencia usados para discretizar las variables numéricas en esta experiencia de laboratorio presentes en la Tabla 1 es posible observar que caen en las categorías: **TSH = Alta**, **T3 = Baja**, **TT4 = Baja** y **FTI = Baja**. Al comparar estas categorías conseguidas del cluster 3 con las reglas que presentan el consecuente *positive* visibles en la Figura 2 es posible observar los antecedentes **TSH = Alta**, **FTI = Baja**, **T3 = Baja** y **TT4 = Baja** lo que concuerda en total exactitud a los antecedentes que representan a la presencia del hipotiroidismo (clase positiva). En resumen, las reglas conseguidas para la clase positiva representan en su totalidad a la agrupación conseguida en el cluster 3 que representa al grupo con hipotiroidismo.

En el caso del cluster 2 (de color verde en la Figura 10) cabe recordar que agrupa a pacientes negativos (Arias and Méndez, 2022), o sea sin enfermedades. Las medias de las hormonas de este cluster son de: TSH = 1.338, T3 = 2.899, TT4 = 156.7, FTI = 138.5. Al ser comparados estos valores a los valores de referencia de la Tabla 1 corresponden a las categorías: **TSH = Normal**, **T3 = Alta**, **TT4 = Alta**, **FTI = Alta**. En el caso de las reglas de asociación conseguidas para la clase negativa, visibles en la Figura 3 es posible observar que se encuentra el valor **Normal** para TSH, FTI, TT4, T3 Y FTI. Al comparar las discretizaciones conseguidas para cada cluster es posible observar que no encajan en totalidad a las conseguidas en las reglas de asociación para la clase negativa, sin embargo es importante notar que en el caso de T3, TT4 y FTI se encuentran muy cerca del límite de haber sido clasificados como normales, con 0.99 nmol/L de diferencia para T3, 5.7 nmol/L para TT4 y 5.8 nmol/L para FTI, tomando esto en consideración y el hecho de que el hipotiroidismo se caracteriza por tener niveles hormonales de TSH alta y FTI baja (Gonzalez, 2021), es que se

considera que el cluster 2 obtiene resultados similares a las reglas de asociación que predicen la clase negativa.

En el caso del cluster 1 (de color rojo en la Figura 10) y como se mencionó en la experiencia anterior (Arias and Méndez, 2022), no agrupa a una clase en específico, presenta mayoritariamente pacientes negativos, pero a su vez concentra a la mayoría de pacientes con hipotiroidismo compensado presentes en la base de datos. Es debido a esto que sus medias hormonales de: T3, TT4, T4U Y FTI son menores, mientras que la de TSH es mayor en comparación al cluster 2 que agrupa a personas sanas. Sus valores medios para cada hormona son de:  $TSH = 3.108$ ,  $T3 = 1.825$ ,  $TT4 = 97.08$ ,  $FTI = 103.6$ , lo que discretizado según los rangos presentados en la Figura 1 se obtiene: **TSH = Normal**, **T3 = Normal**, **TT4 = Normal**, **FTI = Normal**. Resulta interesante que el contenido de este cluster encaja en su totalidad a las reglas de asociación generadas para predecir la clase negativa, esto se puede deber a que como se mencionó, la mayoría de las observaciones dentro del cluster 1, 91 %, para ser exactos corresponden a la clase negativa, habiendo 9 % de clases positivas. Al tomar en cuenta estos resultados, se concluye que es muy probable que si se hubiese realizado el procedimiento de unir todas las clases positivas en una sola y efectuar un clustering con  $K = 2$  (solo dos clusters), las observaciones negativas se hubiesen unido al cluster 2 que agrupa a los negativos, mientras que las observaciones positivas se hubiesen unido al cluster 3 que agrupa a los positivos.

Tomando en consideración la comparación, se hace la conclusión que, para este problema en particular, las reglas de asociación son un mejor método que clustering, debido a que mediante reglas de asociación fue posible el extraer más cantidad de información del problema que con clustering, por ejemplo las relaciones encontradas respecto a las mujeres embarazadas (en reglas pregnant, sección 5.8), la presencia de no cirugía de tiroides en pacientes con hipotiroidismo (en las reglas de clase positiva, sección 5.1), el aumento de la probabilidad de desarrollar hipotiroidismo en mujeres mayores a 40 años (en las reglas de TSH, sección 5.3) y la alteración de hormonas tiroideas en adultos mayores (en reglas T3, sección 5.4). Con clustering no fue posible extraer esta información, solo fue posible lograr conseguir una separación de clases junto a los factores que determinaban cada clase, lo cual también fue obtenido con reglas de asociación.

## 6. Conclusiones

En la presente experiencia de laboratorio se realizó el estudio de la patología de hipotiroidismo mediante reglas de asociación usando la base de datos *Allhypo*.

Los objetivos planteados al inicio de la experiencia se cumplieron, fue posible extraer conocimiento del hipotiroidismo usando reglas de asociación, estas fueron conseguidas usando el paquete *aruleViz* en el software R con el algoritmo *apriori*. Además, la información encontrada fue contrastada a la información adquirida en experiencias de laboratorio anteriores.

Respecto a los resultados obtenidos mediante las reglas de asociación podemos concluir que fueron los esperados, las reglas presentadas para la clase positiva (Sección 5.1) muestran que los niveles hormonales TSH alto, FTI bajo, TT4 bajo y T3 bajo determinan que la persona tenga hipotiroidismo, lo cual concuerda a lo indicado en la literatura (Mayo Clinic, sf). Para el caso de la clase negativa, las reglas presentadas (Sección 5.2) muestran que cuando los niveles hormonales de TSH, FTI, TT4 y T3 se encuentran en el rango normal el paciente no tiene hipotiroidismo, lo cual concuerda con lo que sucede en la realidad.

En cuanto a la comparación con los resultados obtenidos en la experiencia anterior en la cual se usó clustering, cabe destacar que se obtuvieron resultados similares, los clusters agruparon a los pacientes con hipotiroidismo y a los pacientes sin hipotiroidismo, al examinar el cluster 3 (Figura 10) fue posible apreciar que presentaban los mismos parámetros que las reglas de asociación para la predicción de la clase positiva, esto es alto niveles de TSH y bajo nivel en FTI, TT4 y T3. Por otro lado el cluster 2 (Figura 11), agrupa a los pacientes sin hipotiroidismo, al examinarlo es posible observar que se presentan niveles de TSH normales, y niveles altos para T3, TT4 y FTI, sin embargo es importante notar que esta clasificación de niveles altos viene dada por la discretización realizada en la Tabla 1, en donde por un muy poco margen estas hormonas fueron clasificadas como altas. Tomando esto en cuenta consideramos que tanto el cluster 2 como las reglas de asociación presentan resultados similares, agrupando correctamente a pacientes sin hipotiroidismo. Las reglas de asociación para la clase negativa arrojan niveles normales en las hormonas TSH, FTI, TT4 y T3. Como conclusión respecto a la comparación de métodos, se considera que reglas de asociación es un mejor método para este problema específico, al extraer más cantidad de información significativa que clustering como se explicó en la sección 5.9.

Las principales dificultades en esta experiencia fueron: El trabajar con variables numéricas, debido a que el método de reglas de asociación no fue creado para trabajar con variables numéricas, las cuales están presentes en la base de datos, esto fue solucionado mediante la investigación, en donde se llegó a la solución de discretizar los datos, tomando como intervalos los rangos de referencia normales disponibles en la literatura para cada hormona y de etapas para el caso de la edad. Otro problema fue la obtención de un soporte mínimo adecuado para la generación de reglas, este problema también fue resuelto mediante la investigación, en donde se llegó a la solución planteada por una investigación (Lin and Tseng, 2006), en la cual se planteaba la misma problemática del cómo obtener un valor adecuado de soporte mínimo.

Finalmente se considera que se podría mejorar la experiencia, generando nuevas reglas de asociación con algoritmos distintos a priori, con el fin de comparar si estas arrojan un mejor resultado o aportan nueva información que no haya sido obtenida con apriori. Sumado a esto se podrían ocupar nuevos rangos de discretización, debido que estos suelen variar en la literatura dependiendo de los autores, es probable que con distintos rangos de discretización los resultados cambien.



# Bibliografía

- Akmam, E. F., Siswantining, T., Soemartojo, S. M., and Sarwinda, D. (2019). Multiple imputation with predictive mean matching method for numerical missing data. In *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, pages 1–6. IEEE.
- Alice, M. (2015). Imputing missing data with r; mice package. *Data Science Plus*.
- Allison, P. (2015). Imputation by predictive mean matching: Promise & peril. *Statistical Horizons*.
- Amat, J. (2018). Reglas de asociación y algoritmo apriori con r. *cienciadedatos.net*.
- American Thyroid Association (s.f.). Pruebas de función tiroidea. <https://www.thyroid.org/las-pruebas-de-funcion-tiroidea/>.
- Arias, I. and Méndez, C. (2022). Laboratorio 1 – análisis estadístico.
- Brent, G. A. (2022). Thyroid hormone action. *UpToDate*. <https://www.uptodate.com/contents/thyroid-hormone-action>.
- Chacón, M. (2015). Minería de datos capítulo vi: “reglas de asociación”. [Diapositiva de PowerPoint]. Repositorio material curso Análisis de datos.
- Dua, D. and Graff, C. (2017). UCI machine learning repository.
- Galofre, J. (s.f.). Hipotiroidismo. <https://www.cun.es/enfermedades-tratamientos/enfermedades/hipotiroidismo>.
- Gonzalez, A. (2021). Pruebas de función tiroidea. <https://www.cun.es/enfermedades-tratamientos/pruebas-diagnosticas/funcion-tiroidea>.
- Lin, W.-Y. and Tseng, M.-C. (2006). Automated support specification for efficient mining of interesting association rules. *Journal of Information Science*, 32(3):238–250.
- Lodder, P. et al. (2013). To impute or not impute: That’s the question. *Advising on research methods: Selected topics*, pages 1–7.

- López, A. S., Martínez, F. C., and Chávez, J. R. B. (2017). Reglas de asociación en una base de datos del área médica. *Revista Arquitectura e Ingeniería*, 11(2):5.
- Lucas, J. P. (2010). Métodos de clasificación basados en asociación aplicados a sistemas de recomendación. *Salamanca: sn*, pages 30–32.
- Mayo Clinic (s.f.). Hipotiroidismo (tiroides hipoactiva). <https://www.mayoclinic.org/es-es/diseases-conditions/hypothyroidism/diagnosis-treatment/drc-20350289>.
- Ministerio de salud y protección social (s.f.). Ciclo de vida. <https://www.minsalud.gov.co/proteccionsocial/Paginas/cicloVida.aspx>.
- Montero, M. I. M. (2009). Patología tiroidea en el adulto mayor. *Revista Médica de Costa Rica y Centroamérica*, 66(589):245–250.
- Rai, A. (2022). An overview of association rule mining its applications. *UpGrad*.
- Selvi, C. K. and Tamilarasi, A. (2009). An automated association rule mining technique with cumulative support thresholds. *Int. J. Open Problems in Compt. Math*, 2(3).
- Symons, R. and Vining, R. (1985). An evaluation of a fluorescence polarization immunoassay of thyroxin and thyroxin-uptake. *Clinical chemistry*, 31(8):1342–1348.
- Tu otro médico (2020). T4 en la sangre baja. <https://www.tuotromedico.com/parametros/T4-en-sangre-baja.htm>.
- Villanueva, V. (2001). Hipotiroidismo. <https://med.unne.edu.ar/revistas/revista105/Hipotiroidismo>.