



Laboratorio 3: Reglas de asociación

Integrantes: Nicole Reyes
Diego Águila
Curso: Análisis de Datos
Profesor: Max Chacón
Ayudante: Francisco Muñoz

24 de Mayo de 2019

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	2
2.3. Monotonicidad y propiedades de las medidas	3
3. Obtención de reglas	5
3.1. Eliminación de registros vacíos	5
3.2. Eliminación de columnas	5
3.3. Transformación de datos	6
3.4. Función apriori	6
3.5. Reglas obtenidas	6
4. Análisis de los resultados y comparación	9
4.1. Medida de calidad	11
4.2. Relación con experiencia 1	12
4.3. Relación con experiencia 2	12
5. Conclusiones	13
6. Anexos	14
Bibliografía	20

Índice de figuras

- | | | |
|----|---|---|
| 1. | Gráfico de las 10 mejores reglas obtenidas. | 8 |
|----|---|---|

Índice de cuadros

- | | | |
|----|---|---|
| 1. | Tabla correspondiente a las 10 mejores reglas obtenidas | 7 |
| 2. | Tabla correspondiente a las 10 mejores reglas obtenidas | 7 |

1. Introducción

Obtener conocimiento de un conjunto de datos es fundamental al momento de realizar un estudio o análisis con respecto a estos, existen diversas técnicas y estrategias en la minería de datos para lograrlo, sin embargo, la obtención de reglas de asociación es una de las más conocidas y permite establecer relaciones, reglas o inteligencia respecto a lo que puede o no ocurrir.

Esta tercera experiencia de laboratorio consiste en, mediante el proceso de minería de datos y a través de la aplicación de reglas de asociación obtener conocimiento de la base de datos asignada en la primera experiencia ”*Allhypo*” correspondiente al hipotiroidismo.

1.1. Objetivos

1. Profundizar el entendimiento de las reglas de asociación como medida de obtención de conocimiento.
2. Corroborar la teoría vista en clases mediante la aplicación de las reglas de asociación en el lenguaje estadístico R, concretamente en el paquete `rulesViz`.
3. Obtener conocimiento de la base de datos ”*Allhypo*” a través de las reglas de asociación.

A continuación se define un marco teórico con la descripción de las reglas de asociación, sus medidas de calidad y confianza y lo que es la monotonidad de estas. Luego se presentan los resultados de la ejecución del script en R con la obtención de las reglas de asociación para posteriormente realizar un análisis de los resultados y finalizar con las conclusiones respecto a esta experiencia.

2. Marco Teórico

En esta sección se describe en que consisten las reglas de asociación y sus características más importantes.

2.1. Reglas de asociación

Las reglas de asociación describen una relación de asociación entre los elementos de un conjunto de datos, permitiendo establecer conclusiones mediante un conjunto de condiciones a través de reglas. De lo dicho anteriormente se presentan algunos ejemplos:

- Que un cliente compre un producto A, debido a que compra un producto B.
- Si a una persona le gusta un libro A, la probabilidad de que le guste un libro B.
- Si un usuario visita una pagina web A, se puede concluir que visitará una pagina web B.

Una regla de asociación puede expresarse como:

Si A entonces B

Donde A es un conjunto de elementos que condicionan la regla llamados antecedente y B es el conjunto de elementos de la conclusión de la regla llamados consecuente.

2.2. Medidas de calidad y confianza

La confianza de una regla es un porcentaje que indica la frecuencia con la que ocurre una regla, es decir, que tan frecuentemente ocurre el consecuente de una regla en conjunto con el antecedente de esta. La confianza de una regla se puede calcular de la siguiente manera:

$$Conf(A \Rightarrow B) = \frac{Sop(A \Rightarrow B)}{Sop(A)}$$

donde $Sop(A \Rightarrow B)$ corresponde al soporte de la regla, el cual indica la cantidad de veces que se cumple A y B al mismo tiempo en la base de datos.

Las medidas de calidad corresponden a métricas que miden distintas propiedades de una regla, algunas de ellas se definen a continuación:

- **Lift:** Corresponde a la medida de independencia entre el antecedente A y el consecuente B, su valor mínimo es 1 e indica una completa independencia entre A y B y se calcula:

$$Lift = \frac{nConf(A \Rightarrow B)}{Sop(A)}$$

- **Convicción:** También mide la independencia entre el antecedente y el consecuente, sin embargo, en este caso se mantiene la monotonicidad, se calcula de la siguiente manera:

$$Convicción = \frac{n - Sop(B)}{n[1 - Conf(A \Rightarrow B)]}$$

o también mediante probabilidades:

$$Convicción = \frac{P(A)P(\bar{B})}{P(A \cap \bar{B})}$$

- **Laplace:** Es una medida monótona en soporte y confianza y se calcula de la siguiente manera:

$$Laplace(A \Rightarrow B) = \frac{Sop(A \Rightarrow B) + 1}{Sop(A) + k}$$

Donde k es una constante mayor que 1, como $Conf(A \Rightarrow B) = \frac{Sop(A \Rightarrow B)}{Sop(A)}$, laplace se puede escribir:

$$Laplace(A \Rightarrow B) = \frac{Sop(A \Rightarrow B) + 1}{Sop(A \Rightarrow B)/Conf + k}$$

2.3. Monotonidad y propiedades de las medidas

Si se tienen dos especializaciones del antecedente, se generan dos reglas tales que $|A1| < |A2|$ y dos restricciones o medidas $med(A_i)$ con $i=1,2$, asociadas a cada una de las reglas. Se dice que:

La medida es monótona si: $med(A1) \leq med(A2)$.

La medida es anti-monótona si: $med(A1) \geq med(A2)$.(1)

Dado esto, el Soporte corresponde a una medida naturalmente anti-monótona, ya que en la medida que aumentan los atributos del antecedente, el valor de la medida va a

tender a disminuir, ya que habrá menos posibilidades de que ocurran más y más antecedentes al mismo tiempo.

Como la $Conf(A \Rightarrow B) = \frac{Sop(A \Rightarrow B)}{Sop(A)}$, al querer obtener la confianza de un antecedente A y un consecuente B, se asume que el consecuente B no varia, por lo que la monotonidad de la confianza depende de la variación del $Sop(A)$, por lo tanto, si el antecedente aumenta y disminuye el valor del soporte, entonces aumenta el valor de la confianza, en este caso la confianza se vuelve monótona, en caso de que el antecedente A disminuye y por ende aumenta el soporte se provoca una disminución en el valor de la confianza volviéndola anti-monótona.

3. Obtención de reglas

Para obtener las reglas de manera correcta, al igual que en las experiencias anteriores es fundamental elaborar una limpieza a la base de datos. En esta oportunidad la limpieza está compuesta por las siguientes etapas:

3.1. Eliminación de registros vacíos

Para poder realizar una correcta minería de datos nuevamente se vuelve necesario eliminar aquellos pacientes en los cuales hay datos faltantes.

3.2. Eliminación de columnas

Algunas columnas de la base de datos son irrelevantes para la obtención de las reglas puesto que no entregan información con respecto al contexto de los datos y es fundamental eliminarlas ya que en caso contrario entorpecen el proceso de minería de datos volviéndolo ineficaz e ineficiente, debido a que el algoritmo tomaría estas columnas en consideración para la obtención de las reglas. A continuación se mencionan las columnas que se eliminan de la base de datos:

- **TBG:** Se elimina porque esta variable no fue medida y no posee valores en la base de datos.
- **TBG.measured:** Se elimina ya que solo indica si la hormona fue medida o no.
- **T3.measured:** Se elimina ya que solo indica si la hormona fue medida o no.
- **TT4.measured:** Se elimina ya que solo indica si la hormona fue medida o no.
- **T4U.measured:** Se elimina ya que solo indica si la hormona fue medida o no.
- **TSH.measured:** Se elimina ya que solo indica si la hormona fue medida o no.
- **FTI.measured:** Se elimina ya que solo indica si la hormona fue medida o no.
- **referral.source:** Se elimina ya que solo indica la fuente de la que se obtuvo el dato, no entregando información relevante con respecto al hipotiroidismo.

3.3. Transformación de datos

Para el caso de las variables continuas presentes en la base de datos (Age, T3, TT4, T4U, TSH y FTI) se realiza un procedimiento de conversión, en el cual estos datos se transforman a binarios. Para realizar esta transformación se decide utilizar la media de cada elemento, reemplazando aquellos valores que están por sobre la media con un "*True*" y aquellos valores que están por debajo de la media con un "*False*".

3.4. Función apriori

La función utilizada para la obtención de reglas es "*apriori*" y pertenece a la librería "*Arules*". Esta recibe como parámetro, los datos, parámetro objeto de clase para definir soporte y confianza, se recibe un parámetro apariencia con la cual se puede restringir la obtención de reglas, finalmente se tiene un parámetro controlar el cual controla el rendimiento algorítmico de la minería de datos.

3.5. Reglas obtenidas

Para la obtención de las reglas con los datos ya procesados de acuerdo a la limpieza realizada anteriormente. La función que se utiliza para realizar la obtención de las reglas, se especifican los parámetros de la confianza mínima y soporte mínimo los cuales son 0.5 y 0.01 respectivamente. La medida de calidad que se utiliza para este análisis es el *lift* ya que permite comparar la proporción del soporte observado con el teórico, por lo que es mas robusta que la confianza.

Para la obtención de las mejores reglas y que estas entreguen una mejor información acerca de lo que se está estudiando, se aplica una serie de restricciones con respecto a la cantidad de antecedentes y consecuentes, donde el número mínimo es 2 y el número máximo es 5. Entonces, lo que se quiere tener es un único consecuente correspondiente a la variable de *clasificación*, por lo que el número máximo de antecedentes será 4.

A continuación se se muestra una tabla resumen donde se muestran las 10 mejores reglas. Con sus respectivos antecedentes y consecuente:

N° de Regla	Antecedentes	Consecuente
1	on.thyroxine=F, TSH=T, TT4=F	hypothyroid=t
2	on.thyroxine=F, thyroid.surgery=F, TSH=T	hypothyroid=t
3	on.thyroxine=F, TSH=T, T3=F	hypothyroid=t
4	on.thyroxine=F, I131=F, TSH=T	hypothyroid=t
5	on.thyroxine=F, TSH=T, T4U=F	hypothyroid=t
6	on.thyroxine=F, TSH=T, FTI=F	hypothyroid=t
7	on.thyroxine=F, age=F, TSH=T	hypothyroid=t
8	on.thyroxine=F, psych=F, TSH=T	hypothyroid=t
9	on.thyroxine=F, on.antithyroid.med=F, TSH=T	hypothyroid=t
10	on.thyroxine=F, lithium=F, TSH=T	hypothyroid=t

Cuadro 1: Tabla correspondiente a las 10 mejores reglas obtenidas

Se muestra a continuación tabla de resumen de las 10 mejores reglas con sus medidas de calidad.

N° de Regla	Soporte	Confianza	Lift
1	0.067	0.693	8.595
2	0.078	0.692	8.585
3	0.058	0.690	8.568
4	0.078	0.686	8.508
5	0.036	0.685	8.503
6	0.067	0.680	8.437
7	0.035	0.679	8.428
8	0.076	0.678	8.419
9	0.078	0.676	8.395
10	0.078	0.676	8.395

Cuadro 2: Tabla correspondiente a las 10 mejores reglas obtenidas

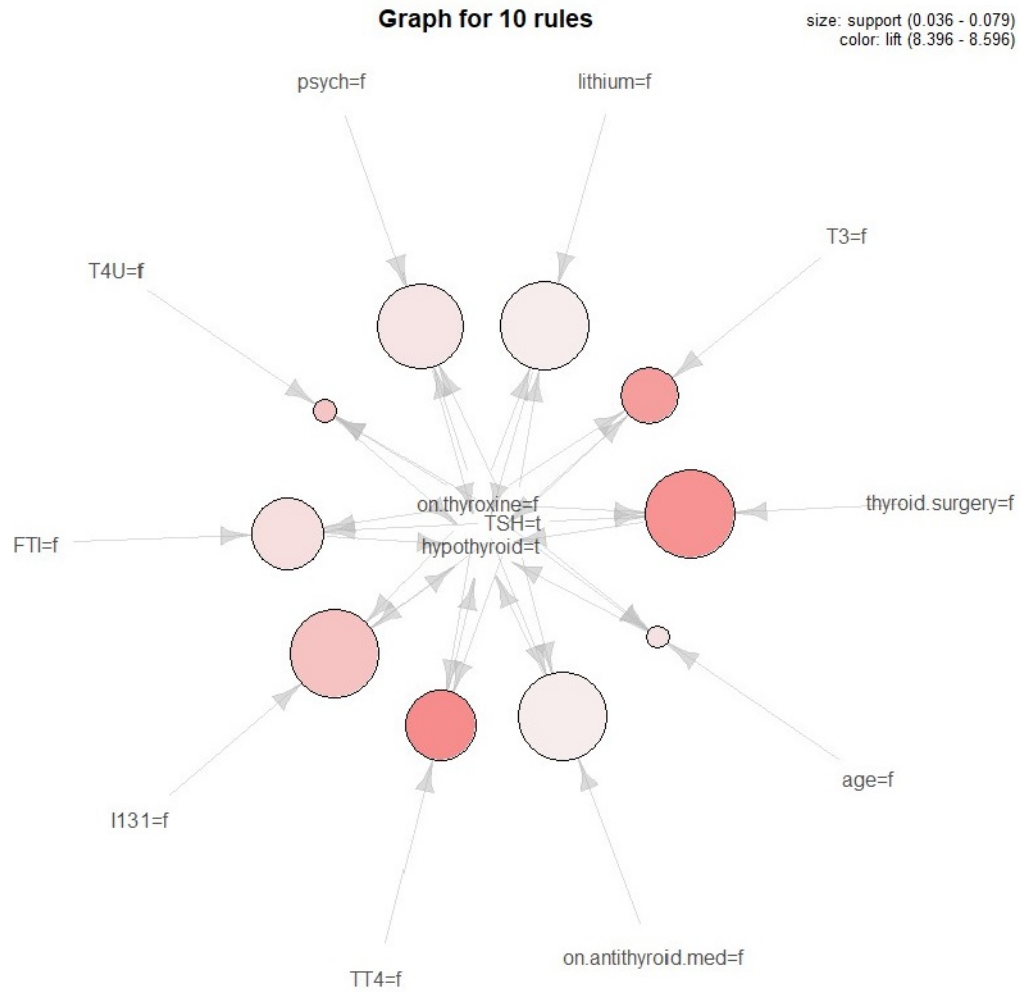


Figura 1: Gráfico de las 10 mejores reglas obtenidas.

La Figura 1 representa un grafo donde cada nodo corresponde a una regla obtenida, el tamaño de este nodo depende del valor obtenido en el soporte de la regla, mientras mayor soporte, más grande será el nodo y viceversa, con respecto al color de los nodos, éste depende del tamaño del lift, mientras mayor sea el lift obtenido en la regla, más colorido estará el nodo, en caso contrario, el nodo tendrá un color más cercano al blanco.

4. Análisis de los resultados y comparación

Una vez obtenidos los resultados de la sección anterior, ahora se procede a realizar un análisis de las reglas. Estas reglas están ordenadas según el valor del lift obtenido por cada una de ellas, en donde las 10 que se muestran son las que tienen mayor valor de esta medida de calidad.

En estas 10 reglas, las variables que están involucradas son *on thyroxine*, *TSH*, *TT4*, *thyroid surgery*, *T3*, *TU4*, *FTI*, *Age*, *psych*, *on antithyroid medication* y *lithium*, donde a partir de la combinación de estas variables se obtiene como consecuente el Hipotiroidismo. A continuación se realiza análisis por cada regla.

1. Regla con antecedente **{on.thyroxine=F, TSH=T, TT4=F}**: De acuerdo a estas variables, se tiene que cuando los pacientes padecen de Hipotiroidismo se tiene una baja secreción de la hormona T4, provocada por un índice alto de TSH, además de que estos pacientes no reciben terapia de reemplazo de Tiroxina, la que es útil al momento de reducir los niveles de TSH en la sangre.
2. Regla con antecedente **{on.thyroxine=F, thyroid.surgery=F, TSH=T}**: De acuerdo a esta regla, se tiene que los que padecen Hipotiroidismo se tiene un nivel alto de TSH en la sangre, la cual es la principal sospecha de padecer de esta enfermedad, además de agregar que el paciente no ha recibido terapia de reemplazo de Tiroxina ni se ha realizado una cirugía a la tiroides en el pasado.
3. Regla con antecedente **{on.thyroxine=F, TSH=T, T3=F}**: De acuerdo a estas variables, se tiene que con un nivel alto de TSH en la sangre, además de no recibir terapia de reemplazo de Tiroxina para reducir estos niveles de TSH por ejemplo, y si se le agrega que los niveles de T3 son bajos, se tiene que el paciente padece de Hipotiroidismo.
4. Regla con antecedente **{on.thyroxine=F, I131=F, TSH=T}**: De acuerdo a esta regla, se tiene que con un nivel alto de secreción de la TSH, además de que el paciente no recibe terapia de reemplazo de la Tiroxina, y si se agrega que el paciente no ha recibido tratamiento de Yodo (I131), el paciente llega a padecer de Hipotiroidismo.

5. Regla con antecedente **{on.thyroxine=F, TSH=T, T4U=F}**: De acuerdo a estas variables se tiene que debido a un alto nivel de TSH en la sangre provocando una baja secreción de la hormona T4, además de que el paciente no ha estado bajo terapia de reemplazo de la Tiroxina, llega a padecer de Hipotiroidismo.
6. Regla con antecedente **{on.thyroxine=F, TSH=T, FTI=F}**: De acuerdo a esta regla se tiene que si se tiene un alto nivel de TSH en la sangre, se explica el bajo nivel de FTI, además de agregar que el paciente no ha recibido terapia de reemplazo de Tiroxina, se tiene como consecuente la condición de Hipotiroidismo.
7. Regla con antecedente **{on.thyroxine=F, age=F, TSH=T}**: De acuerdo a estas variables se tiene que si el paciente no es adulto, presentar niveles altos de TSH en la sangre y además de no recibir terapia de reemplazo de Tiroxina, se tiene como consecuente la condición de Hipotiroidismo.
8. Regla con antecedente **{on.thyroxine=F, psych=F, TSH=T}**: De acuerdo a esta regla se tiene que con altos niveles de TSH en la sangre, además de no recibir terapia de reemplazo de Tiroxina, y agregar que el paciente no sufre de alguna enfermedad psiquiátrica, finalmente como consecuente se tiene la condición de Hipotiroidismo.
9. Regla con antecedente **{on.thyroxine=F, on.antithyroid.med=F, TSH=T}**: De acuerdo a esta regla, se tiene que si se tienen niveles altos de TSH en la sangre, además no recibir medicación antitiroidea ya que es válida para el tratamiento de Hipertiroidismo y no recibir terapia de reemplazo de Tiroxina para reducir estos niveles de TSH, se tiene como consecuente la condición de Hipotiroidismo.
10. Regla con antecedente **{on.thyroxine=F, lithium=F, TSH=T}**: De acuerdo a esta regla, se tiene que con niveles altos de TSH en la sangre, además de no recibir terapia de reemplazo de Tiroxina para disminuir los niveles de TSH, y no recibir tratamiento de Litio, ya que no es recomendable, se tiene como consecuente la condición de Hipotiroidismo.

4.1. Medida de calidad

Para este estudio se utilizaron 3 medidas de calidad, las cuales son el *soporte*, el que es la proporción de observaciones que tienen las variables; la *confianza*, que dice que tan fiable es una regla; y el *lift*, que dado un supuesto de independencia de las observaciones, muestra la proporción del soporte observado con respecto al soporte teórico.

Al realizar un análisis de estas reglas de acuerdo al Cuadro 2 mostrado en la sección anterior, se tienen que estas 10 reglas tienen alto nivel de *lift* y de *confianza*, pero, cabe destacar el bajo nivel de *soporte* que presentan, la explicación es que estas reglas no son tan frecuentes, pero a la vez cada una de estas reglas tiene una gran fiabilidad, con un 95 % de que si se tienen las variables *on thyroxine*, *TSH*, *TT4*, *thyroid surgery*, *T3*, *TU4*, *FTI*, *Age*, *psych*, *on antithyroid medication* y *lithium* como antecedentes, el consecuente será la condición de Hipotiroidismo. Al ser el *lift* claramente mayor a 1, se puede decir que del estudio que se está realizando hay una fuerte asociación entre las variables de este, dando la razón en que las hormonas no son independientes entre ellas, si no que presentan alto grado a relación.

4.2. Relación con experiencia 1

Al realizar estas reglas de asociación, se puede llegar a conclusiones relativamente parecidas con respecto a las variables que afectan directamente al positivo diagnóstico de Hipotiroidismo, como por ejemplo la principal que es la variable que representa la hormona *TSH*, que al ser alta, provoca que los índices de T3 y principalmente de T4 sean bajos. Además de que para la realización de esta experiencia se necesitaba que un pre-procesamiento al igual que la primera experiencia, para así trabajar con los datos adecuados y limpios.

4.3. Relación con experiencia 2

La experiencia 2 consistía en obtener conocimiento de la base de datos mediante la utilización del algoritmo k-means que consiste en generar grupos entre los datos a partir de relaciones que existan entre estos. Con respecto a los resultados obtenidos en la experiencia 2 y en la actual, se observa que en ambos casos se concluye que con una elevada cantidad de hormona TSH mayor es la probabilidad del paciente de padecer hipotiroidismo.

5. Conclusiones

Para obtener las reglas de asociación se realiza en primer lugar una limpieza a la base de datos, en la que se eliminan variables que no entregan información y se transforman las variables continuas en binarias. Una vez obtenida la base de datos trabajable se utiliza el algoritmo apriori para la obtención de las reglas, donde se indica que genere reglas con entre 1 y 3 antecedentes y como consecuente se utiliza la columna de estado que al volverla binaria indica si el paciente tiene o no hipotiroidismo, de este modo, las reglas obtenidas se acotan de forma que indican los antecedentes que pudieran ser causales de esta patología.

Los objetivos de esta experiencia se cumplen, ya que al analizar los resultados se observa que las reglas obtenidas son coherentes con lo obtenido en las experiencias anteriores, entre ellas se enuncian reglas tales como que si el paciente tiene una cantidad elevada de hormona TSH, baja de la hormona TT4 y no se encuentre bajo un tratamiento de tiroxina, entonces con una confianza de 69,3% el paciente puede padecer hipotiroidismo, lo cual se corrobora con lo obtenido en la experiencia 1, otro caso es el de la regla 2, que indica que si el paciente presenta altos niveles de la hormona TSH, no ha tenido cirugía a la tiroides y no se encuentra bajo tratamiento de tiroxina, entonces es probable con una confianza de 69,2% de que el paciente padezca hipotiroidismo.

Para el desarrollo de esta experiencia se presentan algunas dificultades, como por ejemplo, la transformación de las variables continuas a binarias y que criterio utilizar para cumplir este propósito de forma eficaz, otra dificultad que se presenta es identificar la cantidad de antecedentes óptima para obtener las mejores reglas posibles, finalmente se decide utilizar 3 antecedentes ya que al agregar más, el valor de la medida del lift no aumentaba lo suficiente como para que se considerara válido.

6. Anexos

A continuación se presenta el script en R utilizado para el desarrollo de esta experiencia.

```
1 library("arules")
2 library("arulesViz")
3
4 data <- read.csv("C:\\Users\\diego\\Dropbox\\Usach\\ANLISIS DE DATOS\\Lab\\
  allhypo.data",
5               col.names = c("age", "sex", "on.thyroxine", "query.thyroxine",
6               "on.antithyroid.med",
7               "sick", "pregnant", "thyroid.surgery", "I131", "
8               query.hypothyroid",
9               "query.hyperthyroid", "lithium", "goitre", "
10              tumor", "hypopituitary",
11              "psych", "TSH.measured", "TSH", "T3.measured",
12              "T3", "TT4.measured",
13              "TT4", "T4U.measured", "T4U", "FTI.measured", "
14              FTI", "TBG.measured",
15              "TBG", "referral.source", "state"), header=FALSE
16              , sep="," , stringsAsFactors=FALSE)
17
18 # LIMPIEZA DE LA BASE DE DATOS
19
20 # Se limpia la variable de estado del paciente eliminando numeros
21 state <- data$state
22 state <- strsplit(state, ".", fixed = TRUE)
23 state <- lapply(state, '[', 1)
24 state <- unlist(state)
25 data$state <- state
26
27 # Se eliminan los pacientes que contengan algun NA
28 # Sex
29 error = data$sex == '?'
30 data = data[!error,]
31 # Age
```

```

26 error = data$age == '?'
27 data = data[!error,]
28 # On thyroxine
29 error = data$on.thyroxine == '?'
30 data = data[!error,]
31 # Query on antithyroid thyroxine
32 error = data$query.thyroxine == '?'
33 data = data[!error,]
34 # On antithyroid medication
35 error = data$on.antithyroid.med == '?'
36 data = data[!error,]
37 # Sick
38 error = data$sick == '?'
39 data = data[!error,]
40 # Pregnant
41 error = data$pregnant == '?'
42 data = data[!error,]
43 # Thyroid surgery
44 error = data$thyroid.surgery == '?'
45 data = data[!error,]
46 # I131 treatment
47 error = data$I131 == '?'
48 data = data[!error,]
49 # Query hypothyroid
50 error = data$query.hypothyroid == '?'
51 data = data[!error,]
52 # Query hyperthyroid
53 error = data$query.hyperthyroid == '?'
54 data = data[!error,]
55 # Lithium
56 error = data$lithium == '?'
57 data = data[!error,]
58 # Goitre
59 error = data$goitre == '?'
60 data = data[!error,]
61 # Tumor

```

```

62 error = data$tumor == '?'
63 data = data[!error,]
64 # Hypopituitary
65 error = data$hypopituitary == '?'
66 data = data[!error,]
67 # Psych
68 error = data$psych == '?'
69 data = data[!error,]
70 # TSH
71 error = data$TSH == '?'
72 data = data[!error,]
73 # T3
74 error = data$T3 == '?'
75 data = data[!error,]
76 # TT4
77 error = data$TT4 == '?'
78 data = data[!error,]
79 # T4U
80 error = data$T4U == '?'
81 data = data[!error,]
82 # FTI
83 error = data$FTI == '?'
84 data = data[!error,]
85
86 # Se eliminan las variables que indican si se realizo la medicion de la
    hormona
87 data$T3.measured <- NULL
88 data$TSH.measured <- NULL
89 data$TT4.measured <- NULL
90 data$T4U.measured <- NULL
91 data$FTI.measured <- NULL
92 data$TBG.measured <- NULL
93 #Se elimina la variable TBG que no fue medida para ningun paciente
94 data$TBG <- NULL
95 #Se elimina la variable referral.source
96 data$referral.source <- NULL

```

```

97
98 #
99
100 # Una vez limpiada la base de datos, se transforman los datos correspondientes
101
102 # Variables continuas
103 data$T3 <- as.numeric(data$T3)
104 data$age <- as.numeric(data$age)
105 data$TSH <- as.numeric(data$TSH)
106 data$TT4 <- as.numeric(data$TT4)
107 data$T4U <- as.numeric(data$T4U)
108 data$FTI <- as.numeric(data$FTI)
109
110 # Variables binarias
111 data$sex <- as.factor(data$sex)
112 data$sick <- as.factor(data$sick)
113 data$I131 <- as.factor(data$I131)
114 data$state <- as.factor(data$state)
115 data$tumor <- as.factor(data$tumor)
116 data$psych <- as.factor(data$psych)
117 data$goitre <- as.factor(data$goitre)
118 data$lithium <- as.factor(data$lithium)
119 data$pregnant <- as.factor(data$pregnant)
120 data$on.thyroxine <- as.factor(data$on.thyroxine)
121 data$hypopituitary <- as.factor(data$hypopituitary)
122 data$query.thyroxine <- as.factor(data$query.thyroxine)
123 data$thyroid.surgery <- as.factor(data$thyroid.surgery)
124 data$query.hypothyroid <- as.factor(data$query.hypothyroid)
125 data$query.hyperthyroid <- as.factor(data$query.hyperthyroid)
126 data$on.antithyroid.med <- as.factor(data$on.antithyroid.med)
127
128 #Se obtiene la media de los datos continuos para establecerla de manear
    binaria
129 mediaAge <- mean(data$age)

```

```

130 mediaT3 <- mean(data$T3)
131 mediaT4U <- mean(data$T4U)
132 mediaTT4 <- mean(data$TT4)
133 mediaTSH <- mean(data$TSH)
134 mediaFTI <- mean(data$FTI)
135
136 newAge <- integer(length(data[[1]]))
137 newT3 <- integer(length(data[[1]]))
138 newT4U <- integer(length(data[[1]]))
139 newTT4 <- integer(length(data[[1]]))
140 newTSH <- integer(length(data[[1]]))
141 newFTI <- integer(length(data[[1]]))
142
143 #Se transforman los datos continuos a binarios
144 for(i in 1:length(data[[1]])){
145     #si la edad del paciente es mayor que la media, entonces es un
    paciente de edad avanzada
146     if(data$age[i] >= mediaAge){newAge[i] <- "t"}else{newAge[i] <- "f"}
147     #si la cantidad de hormona presente en el paciente es mayor que la
    media se considera como que posee mucho de eso
148     if(data$T3[i] >= mediaT3){newT3[i] <- "t"}else{newT3[i] <- "f"}
149     if(data$TSH[i] >= mediaTSH){newTSH[i] <- "t"}else{newTSH[i] <- "f"}
150     if(data$TT4[i] >= mediaTT4){newTT4[i] <- "t"}else{newTT4[i] <- "f"}
151     if(data$T4U[i] >= mediaT4U){newT4U[i] <- "t"}else{newT4U[i] <- "f"}
152     if(data$FTI[i] >= mediaFTI){newFTI[i] <- "t"}else{newFTI[i] <- "f"}
153 }
154 data$state <- ifelse(data$state %in% c("primary hypothyroid", "secondary
    hypothyroid", "compensated hypothyroid"), "t", "f")
155 data$state <- as.factor(data$state)
156 #Se reemplazan las columnas correspondientes
157 data$age <- as.factor(newAge)
158 data$T3 <- as.factor(newT3)
159 data$TT4 <- as.factor(newTT4)
160 data$T4U <- as.factor(newT4U)
161 data$TSH <- as.factor(newTSH)
162 data$FTI <- as.factor(newFTI)

```

```

163 names(data)[names(data) == "state"] <- "hypothyroid"
164
165 #Se crean las reglas con el mtodo apriori
166 reglas <- apriori(data, parameter = list(minlen=2, support=0.01, confidence
      =0.5, maxlen=4), appearance = list(rhs=c("hypothyroid=t"), default="lhs"))
167
168 #Se muestran las 10 reglas con mejor lift
169 reglas <- sort(reglas, by="lift", decreasing=TRUE)
170 inspect(head(reglas, 10))
171
172 #Se plotean las 10 reglas de mayor lift
173 reglas <- sort(reglas, by="lift", decreasing=TRUE)
174 plot(head(reglas, 10), method="graph", control=list(type="items"))

```

Bibliografía

- [1] Chacón, M. (2018). Reglas de asociación. [Online] <http://www.udesantiagovirtual.cl/moodle2/mod/resource/view.php?id=155128>.
- [IBM] IBM. Reglas de asociación. [Online] https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/nodes_associationrules.html.