



## Laboratorio 3 - Reglas de asociación

Integrantes: Chun-zen Yu  
Matias Pizarro  
Curso: Análisis de Datos  
Sección A-1  
Profesor: Max Chacón Pacheco

8 de Enero de 2021

# Tabla de contenidos

<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos . . . . .	1
<b>2. Marco Teórico</b>	<b>2</b>
2.1. Reglas de asociación . . . . .	2
2.2. Algoritmo Apriori . . . . .	3
2.3. Medidas de calidad . . . . .	4
<b>3. Obtención del reglas</b>	<b>5</b>
3.1. Pre-procesamiento de datos . . . . .	5
3.2. Aplicación de algoritmo Apriori . . . . .	6
<b>4. Análisis de resultado</b>	<b>7</b>
<b>5. Conclusiones</b>	<b>10</b>
<b>Bibliografía</b>	<b>11</b>

# 1. Introducción

Las enfermedades al corazón siempre han sido una gran preocupación para muchas personas, hace años que este tipo de enfermedad ha estado aumentando significativamente, de acuerdo a la World Health Organization (WHO), el numero total de personas que han muerto por enfermedades cardiovasculares llega a los 17,3 millones al año. Sin embargo, un diagnostico durante las primeras etapas de esta enfermedad seguido con el tratamiento adecuado puede salvar una gran cantidad de vidas. Desafortunadamente, el diagnostico correcto de una enfermedad cardiovascular en sus primeras etapas es algo complejo que requiere de un diagnostico médico realizado por expertos en el área, esto es un problema debido a que no todos los centros médicos cuentan con la cantidad adecuada de expertos para todos los pacientes. Es por eso que existe la necesidad de desarrollar un sistema de diagnostico médico de tal forma de poder asistir con este proceso.

En este proyecto se obtendrán reglas de asociación entre los datos, para detectar relaciones de coocurrencia entre atributos, y a través de la confianza y/o soporte asociada a esa relación obtener que atributos y que valor de aquel atributo nos ayuda a detectar o no si hay enfermedad.

## 1.1. Objetivos

- Extraer conocimiento del problema asignado, por medio de las reglas de asociación a través del software R.
- Utilizar el paquete arulesViz.

## 2. Marco Teórico

### 2.1. Reglas de asociación

Las reglas de asociación es una técnica utilizada en la minería de datos y aprendizaje automático, corresponde a una técnica de **aprendizaje no supervisado** el cual nos permiten descubrir combinaciones de atributo-valor que ocurren con frecuencia en un grupo de datos. Estas nos permiten establecer relaciones entre variables cualitativas, esta relación implica concurrencia, no causalidad. ([https://rpubs.com/Cristina\\_Gil/Reglas\\_Asociacion](https://rpubs.com/Cristina_Gil/Reglas_Asociacion))

Este concepto fue basado reglas fuertes descubiertas por *Agrawal et al.* el cual presento un trabajo donde se indicaban las reglas de relación las cuales descubrían las relaciones entre los datos recopilados a gran escala en los sistemas de venta de un supermercado. Por ejemplo se tiene la siguiente regla. (Anonimo, 2010)

$$\{cebolla, vegetales\} \Rightarrow \{carne\} \quad (1)$$

Según esta regla es posible determinar que un consumidor que compra cebolla y vegetales a la vez, es probable que también compre carne. A partir de este tipo de información entregado por las reglas es posible tomar decisiones sobre el marketing como promociones o ubicaciones de los productos dentro del supermercado.

Estas reglas se componen principalmente de 4 componentes:

- Antecedente: Corresponde a la parte de la izquierda de la regla, la cual indica la relación que provoca algo. En este ejemplo corresponde a  $\{cebolla, vegetales\}$ .
- Consecuente: Es la parte de la derecha de la regla, la cual indica la conclusión de la regla. En este ejemplo corresponde a  $\{carne\}$ .
- Soporte: Medida de calidad de la regla.
- Confianza: Medida de calidad de la regla.

## 2.2. Algoritmo Apriori

Existen múltiples algoritmos para la búsqueda de las reglas de asociación, Apriori fue una de los primeros algoritmos desarrollados, este algoritmo consiste en dos etapas:(Hahsler, 2016)

- Identificar todos los itemsets que ocurren con una frecuencia por encima de un determinado limite.
- Convertir esos itemsets frecuentes en reglas de asociación.

A modo de ejemplo se demuestra un caso sencillo de un centro comercial en donde cada fila corresponde a una transacción. Las transacciones hace referencia a todos los productos comprados bajo un misma cesta. en este ejemplo se considera 4 productos (items) distintos.

Transacción
{A, B, C, D}
{A, B, D}
{A, B}
{B, C, D}
{B, C}
{C, D}
{B, D}

Figura 1: Ejemplo de transacciones.

De esta manera el algoritmo Apriori busca entre todas las transacciones los itemsets mas frecuentes. Este proceso no es trivial debido a el gran numero de combinatorias que pueden llegar a existir, una vez identificado esto es relativamente directo generar las reglas de asociación. El algoritmo Apriori realiza una búsqueda exhaustiva por niveles de complejidad. Para reducir el espacio de búsqueda aplica una norma la cual dicta que "si un itemset no es frecuente entonces su superset no es frecuente", aplicado en este ejemplo, si el itemset {A,B} es infrecuente entonces el {A, B, C} Y {A, B, E} también lo son debido a que contienen {A, B} esta regla es conocida como el **principio de monotonicidad**.

## 2.3. Medidas de calidad

Cada regla tiene asociada un valor de **Soporte** y **Confianza**. Estos valores son utilizados para identificar cuales son las reglas mas significativas desde el punto de vista estadístico para el conjunto de datos.

El soporte de un ítem X corresponde al numero de transacciones que contienen X dividido en el total de transacciones. Este valor se representa en la siguiente ecuación.

$$sop(X) = \frac{|X|}{|D|} \quad (2)$$

Mientras que la confianza de una regla se define como La probabilidad que una transacción que contiene los items de X, también contenga los items de Y. Esto se representa según la siguiente ecuación:

$$conf(X \Rightarrow Y) = \frac{sop(XUY)}{sop(X)} \quad (3)$$

Donde la unión de X e Y corresponde a todos los itemsets que contienen X e Y.

### 3. Obtención del reglas

Antes de comenzar a trabajar con los datos es necesario adecuarlos para el estudio que se pretende realizar:

#### 3.1. Pre-procesamiento de datos

Antes de utilizar al algoritmo Apriori es necesario que los datos dentro del conjunto de datos sean variables categóricas. Debido a que el conjunto de datos de Heart Disease consiste de variables numéricas y categóricas es necesario realizar una transformación a las variables numéricas.

Para transformar las variables numéricas se calculo el promedio de cada columna y se estableció un umbral a partir de ese valor, luego se definió un valor para los valores que se encontraban sobre y por debajo del umbral.

```
# Transformar variables numericas a variables categoricas
processed.cleveland.cat <- processed.cleveland
processed.cleveland.cat$age = cut(processed.cleveland$age, breaks = c(-Inf, 54.54, Inf), labels = c("< 54", "> 54"))
processed.cleveland.cat$trestbps = cut(processed.cleveland$trestbps, breaks = c(-Inf, 131.7, Inf), labels = c("Small", "Large"))
processed.cleveland.cat$chol = cut(processed.cleveland$chol, breaks = c(-Inf, 247.4, Inf), labels = c("Small", "Large"))
processed.cleveland.cat$thalach = cut(processed.cleveland$thalach, breaks = c(-Inf, 149.6, Inf), labels = c("Small", "Large"))
processed.cleveland.cat$oldpeak = cut(processed.cleveland$oldpeak, breaks = c(-Inf, 1.05, Inf), labels = c("Small", "Large"))
```

Figura 2: Código R para transformar variables numericas a categoricas

Una vez modificado el conjunto de datos se obtuvo el siguiente resultado:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
1	> 54	male	typical angina	Large	Small	fbs > 120 mg/dl	Estes' criteria	Large	no	Large	downsloping	No major vessel	fixed defect	No present disease
2	> 54	male	asymptomatic	Large	Large	fbs < 120 mg/dl	Estes' criteria	Small	yes	Large	flat	3 major vessels	normal	Present disease
3	> 54	male	asymptomatic	Small	Small	fbs < 120 mg/dl	Estes' criteria	Small	yes	Large	flat	2 major vessels	reversible defect	Present disease
4	< 54	male	non-anginal pain	Small	Large	fbs < 120 mg/dl	normal	Large	no	Large	downsloping	No major vessel	normal	No present disease
5	< 54	female	atypical angina	Small	Small	fbs < 120 mg/dl	Estes' criteria	Large	no	Large	upsloping	No major vessel	normal	No present disease
6	> 54	male	atypical angina	Small	Small	fbs < 120 mg/dl	normal	Large	no	Small	upsloping	No major vessel	normal	No present disease
7	> 54	female	asymptomatic	Large	Large	fbs < 120 mg/dl	Estes' criteria	Large	no	Large	downsloping	2 major vessels	normal	Present disease
8	> 54	female	asymptomatic	Small	Large	fbs < 120 mg/dl	normal	Large	yes	Small	upsloping	No major vessel	normal	No present disease
9	> 54	male	asymptomatic	Small	Large	fbs < 120 mg/dl	Estes' criteria	Small	no	Large	flat	1 major vessel	reversible defect	Present disease
10	< 54	male	asymptomatic	Large	Small	fbs > 120 mg/dl	Estes' criteria	Large	yes	Large	downsloping	No major vessel	reversible defect	Present disease

Figura 3: Dataset despues de transformacion de variables.

## 3.2. Aplicación de algoritmo Apriori

Para realizar la obtención de reglas se utilizo priori con un soporte mínimo de 0.2, es decir todas las reglas con un soporte menor a 0.2 serán eliminados. Además se filtro las reglas que contienen como consecuente si presenta o no una enfermedad cardiaca, esto nos ayuda a saber los principales atributos los cuales indican un caso de tener esta enfermedad. Se realizo en r de la siguiente forma:

```
# Reglas de agrupacion
rules = apriori(
  data = processed.cleveland.cat,
  parameter=list(support = 0.2, minlen = 2, maxlen = 6, target="rules"),
  appearance=list(rhs = c("num=Present disease", "num=No present disease"))
)
```

Figura 4: Código R para transformar variables numericas a categoricas.

Lo que retorno 125 reglas de asociación que serán mostradas y analizadas en el siguiente apartado

```
> inspect(rules)
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{cp=non-anginal pain}	=> {exang=no}	0.2424242	0.8674699	0.2794613	1.2881928	72
[2]	{sex=female}	=> {thal=normal}	0.2693603	0.8333333	0.3232323	1.5091463	80
[3]	{sex=female}	=> {fbs=fbs < 120 mg/dl}	0.2828283	0.8750000	0.3232323	1.0231299	84
[4]	{exang=yes}	=> {cp=asymptomatic}	0.2626263	0.8041237	0.3265993	1.6818644	78
[5]	{exang=yes}	=> {fbs=fbs < 120 mg/dl}	0.2794613	0.8556701	0.3265993	1.0005276	83
[6]	{thal=reversable defect}	=> {sex=male}	0.3367003	0.8695652	0.3872054	1.2848799	100
[7]	{thal=reversable defect}	=> {fbs=fbs < 120 mg/dl}	0.3265993	0.8434783	0.3872054	0.9862718	97
[8]	{oldpeak=Large}	=> {fbs=fbs < 120 mg/dl}	0.3535354	0.8536585	0.4141414	0.9981755	105
[9]	{thalach=Small}	=> {fbs=fbs < 120 mg/dl}	0.3737374	0.8604651	0.4343434	1.0061344	111
[10]	{trestbps=Large}	=> {fbs=fbs < 120 mg/dl}	0.3535354	0.8139535	0.4343434	0.9517488	105
[11]	{chol=Large}	=> {fbs=fbs < 120 mg/dl}	0.3872054	0.8518519	0.4545455	0.9960630	115
[12]	{slope=flat}	=> {fbs=fbs < 120 mg/dl}	0.4040404	0.8759124	0.4612795	1.0241968	120
[13]	{num=Present disease}	=> {sex=male}	0.3771044	0.8175182	0.4612795	1.2079747	112
[14]	{num=Present disease}	=> {fbs=fbs < 120 mg/dl}	0.3939394	0.8540146	0.4612795	0.9985919	117
[15]	{age<= 54}	=> {fbs=fbs < 120 mg/dl}	0.4141414	0.8913043	0.4646465	1.0421945	123

Figura 5: 15 de 125 reglas obtenidas.

Luego para identificar las reglas mas importantes para el conjunto de datos heart disease se aplico un filtro de modo que el consecuente de la regla sea la variable dependiente *num* la cual es la variable que indica si el paciente padece de alguna enfermedad cardiaca. Y además se aplico un ordenamiento decreciente según el soporte y la confianza de las reglas. Luego se escogió las 10 primeras reglas obtenidas para analizar el comportamiento del conjunto de datos.



## 4. Análisis de resultado

Como se hablo en el capitulo anterior, en este se revisara los resultados obtenidos en este caso reglas de asociación obtenidas. Las cuales como se mostrara a continuación fueron ordenadas de dos formas distintas, una es descendente por confianza y descendente por soporte:

```
> inspect(sort(x = rules, decreasing = TRUE, by = "confidence"))
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{age<= 54,thalach=Large,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2020202	0.9836066	0.2053872	1.825820	60
[2]	{sex=female,num=No present disease}	=> {thal=normal}	0.2323232	0.9718310	0.2390572	1.759962	69
[3]	{age<= 54,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2323232	0.9718310	0.2390572	1.803961	69
[4]	{age<= 54,fbs=fbs < 120 mg/dl,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2188552	0.9701493	0.2255892	1.800840	65
[5]	{sex=female,fbs=fbs < 120 mg/dl,num=No present disease}	=> {thal=normal}	0.2121212	0.9692308	0.2188552	1.755253	63
[6]	{sex=female,exang=no,num=No present disease}	=> {thal=normal}	0.2053872	0.9682540	0.2121212	1.753484	61
[7]	{age<= 54,exang=no,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2053872	0.9682540	0.2121212	1.797321	61
[8]	{trestbps=Small,exang=no,slope=upsloping}	=> {oldpeak=Small}	0.2222222	0.9565217	0.2323232	1.632684	66
[9]	{cp=asymptomatic,ca=No major vessel}	=> {fbs=fbs < 120 mg/dl}	0.2087542	0.9538462	0.2188552	1.115324	62
[10]	{age<= 54,thalach=Large,thal=normal}	=> {num=No present disease}	0.2323232	0.9452055	0.2457912	1.754538	69

Figura 6: Reglas de agrupación ordenadas por confianza.

```
> inspect(sort(x = rules, decreasing = TRUE, by = "support"))
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{exang=no}	=> {fbs=fbs < 120 mg/dl}	0.5757576	0.8550000	0.6734007	0.9997441	171
[2]	{sex=male}	=> {fbs=fbs < 120 mg/dl}	0.5723906	0.8457711	0.6767677	0.9889529	170
[3]	{ca=No major vessel}	=> {fbs=fbs < 120 mg/dl}	0.5218855	0.8908046	0.5858586	1.0416101	155
[4]	{trestbps=Small}	=> {fbs=fbs < 120 mg/dl}	0.5016835	0.8869048	0.5656566	1.0370501	149
[5]	{oldpeak=Small}	=> {fbs=fbs < 120 mg/dl}	0.5016835	0.8563218	0.5858586	1.0012897	149
[6]	{thal=normal}	=> {fbs=fbs < 120 mg/dl}	0.4848485	0.8780488	0.5521886	1.0266948	144
[7]	{thalach=Large}	=> {fbs=fbs < 120 mg/dl}	0.4814815	0.8511905	0.5656566	0.9952897	143
[8]	{thalach=Large}	=> {exang=no}	0.4713805	0.8333333	0.5656566	1.2375000	140
[9]	{chol=Small}	=> {fbs=fbs < 120 mg/dl}	0.4680135	0.8580247	0.5454545	1.0032808	139
[10]	{num=No present disease}	=> {exang=no}	0.4612795	0.8562500	0.5387205	1.2715312	137

Figura 7: Reglas de agrupación ordenadas por soporte.

En este conjunto de datos la variable dependiente corresponde a *num* la cual indica si el paciente tiene algún tipo de enfermedad cardiaca. Dado esto, se aplico un filtro sobre las reglas para identificar las reglas que tengan como consecuente la variable *num*.

Primero se realizo un ordenamiento decreciente con la confianza, como se demuestra en la figura 8 y figura 9 donde se muestra las primeras 10 reglas una vez aplicado el filtro.

```
> inspect(sort(x = rules, decreasing = TRUE, by = "confidence"))
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{age=< 54,thalach=Large,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2020202	0.9836066	0.2053872	1.825820	60
[2]	{age=< 54,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2323232	0.9718310	0.2390572	1.803961	69
[3]	{age=< 54,fbs=fbs < 120 mg/dl,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2188552	0.9701493	0.2255892	1.800840	65
[4]	{age=< 54,exang=no,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2053872	0.9682540	0.2121212	1.797321	61
[5]	{age=< 54,thalach=Large,thal=normal}	=> {num=No present disease}	0.2323232	0.9452055	0.2457912	1.754538	69
[6]	{age=< 54,thalach=Large,exang=no,thal=normal}	=> {num=No present disease}	0.2087542	0.9393939	0.2222222	1.743750	62
[7]	{age=< 54,fbs=fbs < 120 mg/dl,thalach=Large,thal=normal}	=> {num=No present disease}	0.2087542	0.9393939	0.2222222	1.743750	62
[8]	{thalach=Large,oldpeak=Small,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2087542	0.9393939	0.2222222	1.743750	62
[9]	{trestbps=Small,exang=no,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2053872	0.9384615	0.2188552	1.742019	61
[10]	{trestbps=Small,ca=No major vessel,thal=normal}	=> {num=No present disease}	0.2255892	0.9305556	0.2424242	1.727344	67

Figura 8: Reglas de agrupación con consecuente de enfermedad cardiaca ordenadas por confianza.

```
> inspect(sort(x = rules, decreasing = TRUE, by = "confidence"))
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{cp=asymptomatic,thal=reversible defect}	=> {num=Present disease}	0.2356902	0.9090909	0.2592593	1.970803	70
[2]	{cp=asymptomatic,exang=yes}	=> {num=Present disease}	0.2289562	0.8717949	0.2626263	1.889949	68
[3]	{thalach=Small,exang=yes}	=> {num=Present disease}	0.2020202	0.8695652	0.2323232	1.885116	60
[4]	{sex=male,cp=asymptomatic,thalach=Small}	=> {num=Present disease}	0.2020202	0.8695652	0.2323232	1.885116	60
[5]	{cp=asymptomatic,oldpeak=Large}	=> {num=Present disease}	0.2188552	0.8666667	0.2525253	1.878832	65
[6]	{thalach=Small,oldpeak=Large}	=> {num=Present disease}	0.2188552	0.8227848	0.2659933	1.783701	65
[7]	{cp=asymptomatic,thalach=Small}	=> {num=Present disease}	0.2457912	0.8202247	0.2996633	1.778151	73
[8]	{cp=asymptomatic,fbs=fbs < 120 mg/dl,thalach=Small}	=> {num=Present disease}	0.2121212	0.8076923	0.2626263	1.750983	63
[9]	{cp=asymptomatic,slope=flat}	=> {num=Present disease}	0.2222222	0.8048780	0.2760943	1.744882	66
[10]	{sex=male,exang=yes}	=> {num=Present disease}	0.2020202	0.8000000	0.2525253	1.734307	60

Figura 9: Reglas de agrupación con consecuente de enfermedad cardiaca ordenadas por confianza.

Como se puede observar dentro de las reglas ordenadas por confianza se puede observar que las variable  $age = j54$ ,  $ca = No\ major\ vessel$ ,  $thal = normal$  y  $thalach = Large$  se encuentran presentes en varias ocasiones, esto indica que una persona que tenga estas condiciones es mas probable que no presente una enfermedad cardiaca. Mientras que para los casos donde si hubo una enfermedad cardiaca las variables  $cp = asymptomatic$ ,  $exang = yes$  y  $oldpeak = large$  estuvieron mas presentes. Sin embargo el soporte de estas reglas no superaron el 0.25, lo que quiere decir que pacientes que presenten estas condiciones no son comunes.

Por el otro lado, también se realizo un ordenamiento decreciente con el soporte, como se demuestra en la figura 10 y la figura 11, donde se muestra las primeras 10 de las reglas obtenidas.

```

> inspect(sort(x = rules, decreasing = TRUE, by = "support"))
  lhs                                     rhs      support  confidence coverage lift  count
[1] {exang=no,thal=normal}                => {num=No present disease} 0.3737374 0.8345865 0.4478114 1.549201 111
[2] {exang=no,ca=No major vessel}          => {num=No present disease} 0.3737374 0.8473282 0.4410774 1.572853 111
[3] {ca=No major vessel,thal=normal}        => {num=No present disease} 0.3434343 0.8869565 0.3872054 1.646413 102
[4] {thalach=Large,ca=No major vessel}      => {num=No present disease} 0.3333333 0.8461538 0.3939394 1.570673 99
[5] {thalach=Large,thal=normal}             => {num=No present disease} 0.3299663 0.8448276 0.3905724 1.568211 98
[6] {fbs=fbs < 120 mg/dl,exang=no,thal=normal} => {num=No present disease} 0.3265993 0.8362069 0.3905724 1.552209 97
[7] {fbs=fbs < 120 mg/dl,exang=no,ca=No major vessel} => {num=No present disease} 0.3265993 0.8290598 0.3939394 1.538942 97
[8] {oldpeak=Small,thal=normal}             => {num=No present disease} 0.3232323 0.8205128 0.3939394 1.523077 96
[9] {oldpeak=Small,ca=No major vessel}      => {num=No present disease} 0.3164983 0.8173913 0.3872054 1.517283 94
[10] {fbs=fbs < 120 mg/dl,ca=No major vessel,thal=normal} => {num=No present disease} 0.3097643 0.8761905 0.3535354 1.626429 92

```

Figura 10: Reglas de agrupación con consecuente de enfermedad cardiaca ordenadas por soporte.

```

> inspect(sort(x = rules, decreasing = TRUE, by = "support"))
  lhs                                     rhs      support  confidence coverage lift  count
[1] {cp=asymptomatic,thalach=Small}          => {num=Present disease} 0.2457912 0.8202247 0.2996633 1.778151 73
[2] {cp=asymptomatic,thal=reversible defect} => {num=Present disease} 0.2356902 0.9090909 0.2592593 1.970803 70
[3] {cp=asymptomatic,exang=yes}              => {num=Present disease} 0.2289562 0.8717949 0.2626263 1.889949 68
[4] {cp=asymptomatic,slope=flat}            => {num=Present disease} 0.2222222 0.8048780 0.2760943 1.744882 66
[5] {thalach=Small,oldpeak=Large}            => {num=Present disease} 0.2188552 0.8227848 0.2659933 1.783701 65
[6] {cp=asymptomatic,oldpeak=Large}          => {num=Present disease} 0.2188552 0.8666667 0.2525253 1.878832 65
[7] {cp=asymptomatic,fbs=fbs < 120 mg/dl,thalach=Small} => {num=Present disease} 0.2121212 0.8076923 0.2626263 1.750983 63
[8] {thalach=Small,exang=yes}                => {num=Present disease} 0.2020202 0.8695652 0.2323232 1.885116 60
[9] {sex=male,exang=yes}                    => {num=Present disease} 0.2020202 0.8000000 0.2525253 1.734307 60
[10] {sex=male,cp=asymptomatic,thalach=Small} => {num=Present disease} 0.2020202 0.8695652 0.2323232 1.885116 60

```

Figura 11: Reglas de agrupación con consecuente de enfermedad cardiaca ordenadas por soporte.

En las reglas ordenadas por soporte donde los pacientes no presentaban una enfermedad cardiaca, el soporte llego a tener un 0.37 lo que no es muy alto pero aun así indica los itemset mas comunes entre los pacientes, las variables *exang = no*, *ca = no major vessel* y *thal = normal* se encuentran presentes en múltiples ocasiones, estas variables también se encontraban presentes en el ordenamiento por confianza, lo que pueden demostrar que pueden ser un buen identificador para verificar que el paciente no presenta una enfermedad cardiaca.

Por el otro lado las reglas ordenadas por soporte donde los paciente si presentaban una enfermedad cardiaca se obtuvo los mismo resultados debido a que solo consistía en 10 reglas.

## 5. Conclusiones

Este laboratorio se logro desarrollar de manera correcta, se obtuvieron las reglas de asociación mas importantes para el conjunto de datos según la confianza y soporte de cada una de estas. Gracias a la obtención de estas reglas se pudo realizar un estudio mas profundo acerca de las variables del conjunto de datos, gracias al filtro utilizado donde se identifico las reglas cuyo consecuente correspondía a la variable la cual indicaba si el paciente presenta una enfermedad cardiaca, es posible identificar que variables son las mas importantes al momento de intentar predecir si el paciente podrá padecer de algún tipo de enfermedad cardiaca en el futuro. Sin embargo la mayoría de las obtenidas no superaban el 30 % de soporte pero tenían una confianza de 80 %+, esto indica que las reglas no son muy comunes dentro del conjunto de dato, pero los casos los cuales si se presentan los antecedentes el consecuente es muy probable.

Al analizar los resultados obtenidos y compararlos con los resultados obtenidos en el laboratorio anterior donde se construyo un cluster, se puede notar que los resultados obtenidos en esta experiencia son mas concretos, debido a que la técnica de reglas de asociación es mas adecuada para este conjunto de datos. Esto se debe a que las variables de este conjunto de datos corresponden en su mayoría a variables categóricas por lo que no hay una alteración significativa de los datos.

Con respecto a los objetivos de este laboratorio se cumplieron en su totalidad, ya que logro extraer conocimientos del problema, a través del uso del apriori para la obtención de reglas de asociación, que nos permite inferir sobre relaciones entre atributos. Lo anterior fue implementado en R, haciendo uso del paquete “ arules”

# Bibliografía

- Amat, J. (2018). Reglas de asociación y algoritmo apriori con r. [Online] [https://www.cienciadedatos.net/documentos/43\\_reglas\\_de\\_asociacion](https://www.cienciadedatos.net/documentos/43_reglas_de_asociacion).
- Anonimo (2010). Arules. [Online] [https://es.wikipedia.org/wiki/Reglas\\_de\\_asociaci%C3%B3n](https://es.wikipedia.org/wiki/Reglas_de_asociaci%C3%B3n).
- Garg, A. (2018). Complete guide to association rules (1/2). [Online] <https://towardsdatascience.com/association-rules-2-aa9a77241654>.
- Hahsler, M. (2015). Arules. [Online] <https://www.rdocumentation.org/packages/arules/versions/1.6-6>.
- Hahsler, M. (2016). Apriori. [Online] <https://www.rdocumentation.org/packages/arules/versions/1.6-6/topics/apriori>.
- Martínez, C. G. (2020). Reglas de asociación. [Online] [https://rpubs.com/Cristina\\_Gil/Reglas\\_Asociacion](https://rpubs.com/Cristina_Gil/Reglas_Asociacion).