



UNIVERSIDAD DE GRANADA

PRÁCTICA 4: REGLAS DE ASOCIACIÓN

Autor

Juan Manuel Castillo Nievas



MÁSTER PROFESIONAL EN INGENIERÍA INFORMÁTICA

Granada, 17 de diciembre de 2020

Índice

1. Introducción	2
2. Preprocesamiento de datos	3
2.1. Matriz de correlación	3
2.2. Discretización de Age	4
2.3. Discretización de CCAvg	5
2.4. Discretización de Income	6
2.5. Discretización de Mortgage	7
2.6. Convertir Education en atributo nominal	8
2.7. Convertir Personal Loan en atributo nominal	9
2.8. Convertir Securities Account en atributo nominal	9
2.9. Convertir CD Account en atributo nominal	10
2.10. Convertir Online en atributo nominal	10
2.11. Convertir CreditCard en atributo nominal	11
2.12. Atributos no interesantes	11
3. Reglas de asociación	12
3.1. Reglas que más veces ocurren	15
3.2. Reglas cuyo RHS = Securities Account	16
3.3. Reglas cuyo RHS = CD Account	18
3.4. Reglas cuyo LHS = Income y Mortgage	20
3.5. Reglas cuyo LHS = Age y Mortgage	22

1. Introducción

UniversalBank desarrolló una campaña para convertir sus clientes pasivo en clientes activo. Para diseñar campañas más efectivas, se propone responder a la siguiente pregunta: “Hay varios productos/servicios que el banco ofrece como tales como cuentas de valores (securities accounts), certificados de depósitos, servicios de banca en línea o tarjetas de crédito. ¿Se puede detectar alguna relación entre estos productos para encontrar oportunidades de venta cruzada?”

Para responder a esta pregunta se deben utilizar reglas de asociación para encontrar la relación entre los productos.

2. Preprocesamiento de datos

2.1. Matriz de correlación

Se ha obtenido la matriz de correlación para eliminar posibles atributos que tengan una muy fuerte relación. En la Figura 1 se muestra la matriz de correlación. Tal y como se puede ver, la **edad** y la **experiencia laboral** están ligadas muy fuertemente, con lo cual se va a **eliminar la variable Experience**.

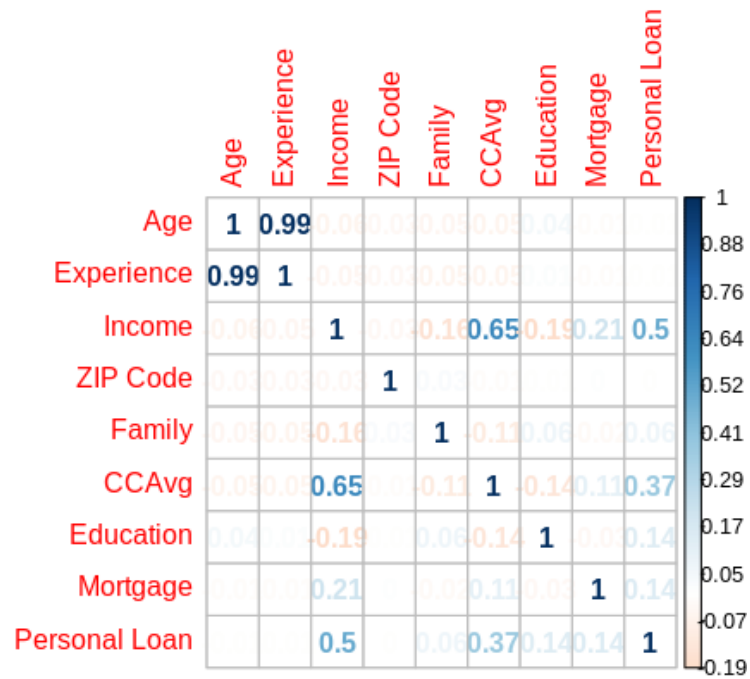


Figura 1: Matriz de correlación

2.2. Discretización de Age

La variable **Age** describe la edad de los clientes. Se ha propuesto un agrupamiento de los datos por igualdad de frecuencia. Los grupos se muestran en la Figura 2. En la Listing 1 se muestra el código en R usado para ello.

```
> table(discretize(prestamo$Age, method="frequency", breaks = 4))  
  
[23,35) [35,45) [45,55) [55,67]  
    1123     1269     1267     1341
```

Figura 2: Discretización de **Age**

```
1  # Discretización de Age  
2  table(discretize(prestamo$Age, method="frequency", breaks = 4))  
3  prestamo <- discretizeDF(prestamo, methods = list(  
4    Age = list(method = "frequency", breaks = 4,  
5              labels = c("23-34", "35-44", "45-54", "55-67"))  
6    ),  
7    default = list(method = "none")  
8  )
```

Listing 1: Discretización de **Age** en R

2.3. Discretización de CCAvg

La variable **CCAvg** el gasto que hacen los clientes de media a lo largo de un mes. Se ha propuesto un agrupamiento de los datos por intervalos. Los grupos se muestran en la Figura 3. En la Listing 2 se muestra el código en R usado para ello.

```
> table(discretize(prestamo$CCAvg, method="interval", breaks = 5))  
  
[0,2) [2,4) [4,6) [6,8) [8,10]  
3059   1358   316   216    51
```

Figura 3: Discretización de **CCAvg**

```
1  # Discretización de CCAvg  
2  table(discretize(prestamo$CCAvg, method="interval", breaks = 5))  
3  prestamo <- discretizeDF(prestamo, methods = list(  
4    CCAvg = list(method = "interval", breaks = 5,  
5                labels = c("[0-2)", "[2-4)", "[4-6)", "[6-8)", "[8-10)"))  
6    ),  
7    default = list(method = "none")  
8  )
```

Listing 2: Discretización de **CCAvg** en R

2.4. Discretización de Income

La variable **Income** describe el salario anual del cliente. Se ha propuesto un agrupamiento de los datos por igualdad de frecuencia. Los grupos se muestran en la Figura 4. En la Listing 3 se muestra el código en R usado para ello.

```
> table(discretize(prestamo$CAvg, method="interval", breaks = 5))
```

[0,2)	[2,4)	[4,6)	[6,8)	[8,10]
3059	1358	316	216	51

Figura 4: Discretización de **Income**

```
1  # Discretización de Income
2  table(discretize(prestamo$Income, method="frequency", breaks = 5))
3  prestamo <- discretizeDF(prestamo, methods = list(
4    Income = list(method = "frequency", breaks = 5,
5                  labels = c("[8-33)", "[33-52)", "[52-78)", "[78-113)", "[113-224]"))
6    ),
7    default = list(method = "none")
8  )
```

Listing 3: Discretización de **Income** en R

2.5. Discretización de Mortgage

La variable **Mortgage** describe el valor de la hipoteca de la casa del cliente. Se ha propuesto un agrupamiento dividiendo los datos en:

- **No tiene hipoteca:** si el valor es 0
- **Hipoteca normal:** si el valor está entre (0,250)
- **Hipoteca alta:** si el valor es mayor de 250

El histograma de esta variable se encuentra en la Figura 5, y es el que se ha usado para proponer la agrupación. En la Listing 4 se muestra el código en R usado para ello.

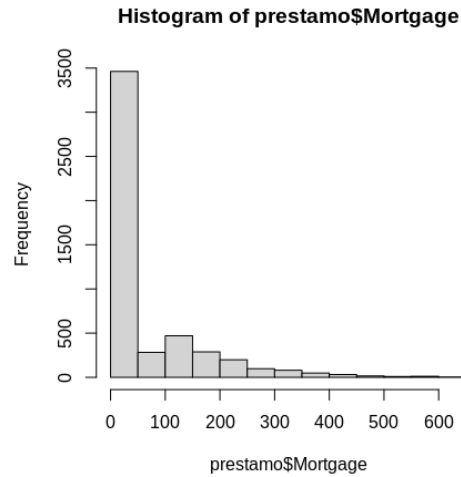


Figura 5: Histograma de **Mortgage**

```
1 # Discretización de la variable Mortgage
2 hist(prestamo$Mortgage)
3 prestamo$Mortgage <- cut(prestamo$Mortgage,
4                           breaks = c(-Inf, 1, 250, +Inf),
5                           labels = c("No tiene hipoteca", "Hipoteca normal",
6                                       "Hipoteca alta"), right = FALSE)
```

Listing 4: Discretización de **Mortgage** en R

2.6. Convertir Education en atributo nominal

Este atributo describe el nivel de educación del cliente. Hay 4 niveles:

- **Education level:** si el valor es 1
- **Undergrad:** si el valor es 2
- **Graduate:** si el valor es 3
- **Advanced:** si el valor es 4

El código usado para esto se muestra en la Listing 5.

```
1  # Pasar Education a nominal
2  prestamo$Education <- cut(prestamo$Education,
3                           breaks = c(1, 2, 3, 4, +Inf),
4                           labels = c("Education level", "Undergrad",
5                                     "Graduate", "Advanced"),
6                           right = FALSE)
```

Listing 5: Convertir **Education** en nominal

2.7. Convertir Personal Loan en atributo nominal

Este atributo describe si el cliente aceptó un préstamo personal en la última campaña. Hay 2 niveles:

- **No aceptó préstamo personal:** si el valor es 0
- **Aceptó préstamo personal:** si el valor es 1

El código usado para esto se muestra en la Listing 7.

```
1  # Pasar Personal Loan a nominal
2  prestamo$`Personal Loan` <- cut(prestamo$`Personal Loan`,
3                                breaks = c(0, 1, +Inf),
4                                labels = c("No acepto prestamo personal"
5                                           , "Acepto prestamo personal"), right = FALSE)
```

Listing 6: Convertir **Personal Loan** en nominal

2.8. Convertir Securities Account en atributo nominal

Este atributo describe si el cliente tiene cuentas de seguridad en el banco. Hay 2 niveles:

- **No tiene cuentas de seguridad:** si el valor es 0
- **Tiene cuentas de seguridad:** si el valor es 1

El código usado para esto se muestra en la Listing 7.

```
1  # Pasar Securities account a nominal
2  prestamo$`Securities Account` <- cut(prestamo$`Securities Account`,
3                                     breaks = c(0, 1, +Inf),
4                                     labels = c("No tiene cuentas de seguridad",
5                                                "Tiene cuentas de seguridad"), right = FALSE)
```

Listing 7: Convertir **Securities Account** en nominal

2.9. Convertir CD Account en atributo nominal

Este atributo describe si el cliente tiene una cuenta de certificado de depósito en el banco. Hay 2 niveles:

- **No tiene cuenta CD:** si el valor es 0
- **Tiene cuenta CD:** si el valor es 1

El código usado para esto se muestra en la Listing 8.

```
1  # Pasar CD account a nominal
2  prestamo$`CD Account` <- cut(prestamo$`CD Account`,
3                               breaks = c(0, 1, +Inf),
4                               labels = c("No tiene cuenta CD", "Tiene cuenta CD"),
5                               right = FALSE)
```

Listing 8: Convertir **CD Account** en nominal

2.10. Convertir Online en atributo nominal

Este atributo describe si el cliente usa la banca online. Hay 2 niveles:

- **No usa online:** si el valor es 0
- **Usa online:** si el valor es 1

El código usado para esto se muestra en la Listing 9.

```
1  # Pasar Online a nominal
2  prestamo$Online <- cut(prestamo$Online,
3                          breaks = c(0, 1, +Inf),
4                          labels = c("No usa online", "Usa online"),
5                          right = FALSE)
```

Listing 9: Convertir **Online** en nominal

2.11. Convertir CreditCard en atributo nominal

Este atributo describe si el cliente usa una tarjeta de crédito del banco. Hay 2 niveles:

- **No usa tarjeta:** si el valor es 0
- **Usa tarjeta:** si el valor es 1

El código usado para esto se muestra en la Listing 10.

```
1  # Pasar CreditCard a nominal
2  prestamo$CreditCard <- cut(prestamo$CreditCard,
3                             breaks = c(0, 1, +Inf),
4                             labels = c("No usa tarjeta", "Usa tarjeta"),
5                             right = FALSE)
```

Listing 10: Convertir **CreditCard** en nominal

2.12. Atributos no interesantes

En este problema se ha considerado que el atributo **ZIP Code** no resulta interesante para este tipo de problema, ya que no es relevante el código postal para saber si un cliente es potencial o no.

3. Reglas de asociación

Se ha usado la librería **arules** para utilizar el algoritmo **apriori** y de esa forma identificar los objetos más frecuentes y crear reglas de asociación.

Para utilizar este algoritmo, primero hay que convertir nuestro dataset en transacciones. En el código que se muestra en la Listing 11 se puede ver el código usado para esto. También se ha hecho una gráfica que muestra los objetos más frecuentes de cada atributo.

```
1 library(arules)
2 prestamo_trans = as(prestamo, "transactions")
3 summary(prestamo_trans)
4 itemFrequencyPlot(prestamo_trans, topN = 11)
```

Listing 11: Convertir el dataset a transacciones

En la Figura 7 se muestra el resumen de las transacciones. Se puede ver que lo más frecuente es que los clientes **no tengan cuenta de certificado de depósito**, seguido de los clientes que **no tienen cuentas de seguridad**. También es frecuente que los usuarios **no tengan hipoteca**.

```
transactions as itemMatrix in sparse format with
5000 rows (elements/itemsets/transactions) and
35 columns (items) and a density of 0.3142857

most frequent items:
      CD Account=No tiene cuenta CD      Personal Loan=No acepto prestamo personal
                                4698                                4520
Securities Account=No tiene cuentas de seguridad      CreditCard=No usa tarjeta
                                4478                                3530
      Mortgage=No tiene hipoteca      (Other)
                                3462                                34312
```

Figura 6: Resumen de las transacciones

En la Figura 7 se muestran los valores más frecuentes de cada atributo.

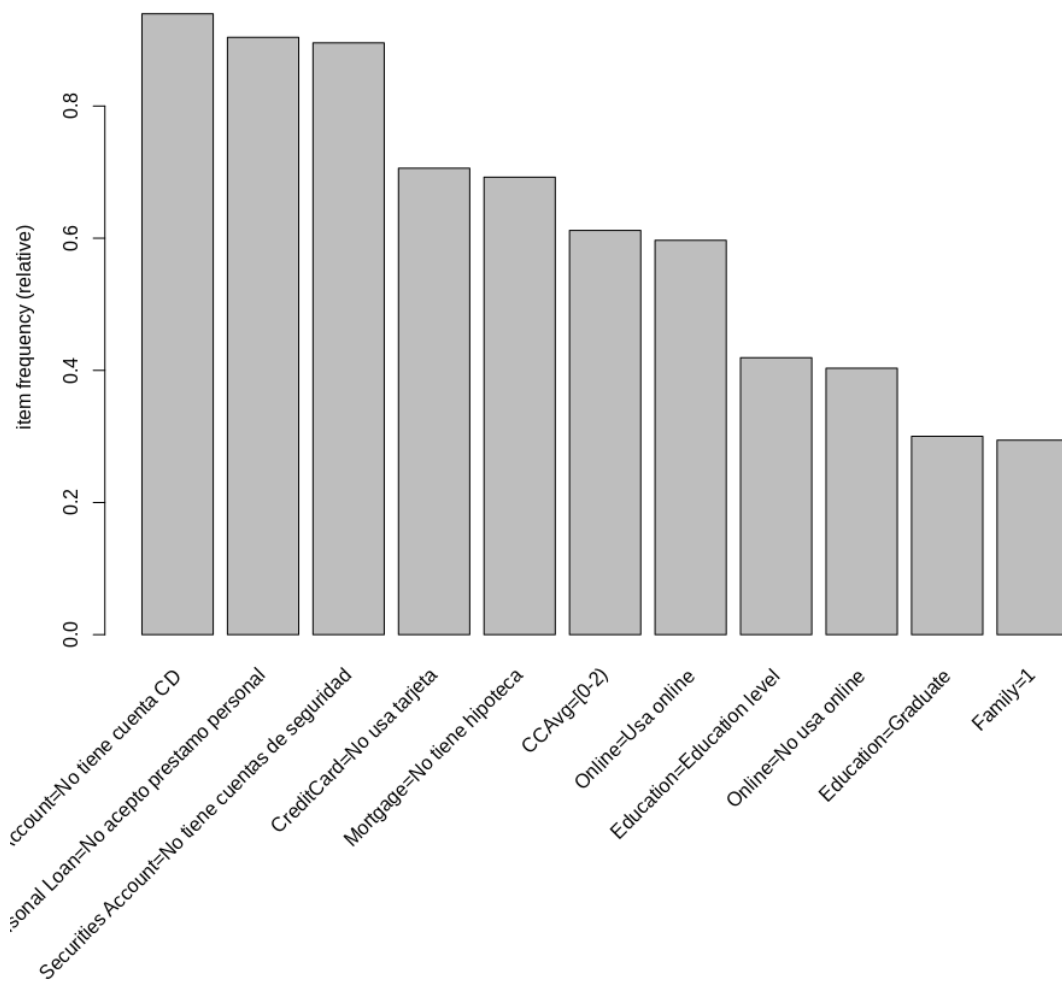


Figura 7: Valor más frecuente de cada atributo

En el código que se muestra en la Listing 12 se ve cómo se ha aplicado el **algoritmo apriori**. Como parámetro, se ha propuesto que las reglas de asociación tenga una confianza de más del 95 %, para asegurarnos de que esas reglas son verdad en casi el 100 % de los casos.

```
1 rules = apriori(data = prestamo_trans, parameter = list(confidence = 0.95))
```

Listing 12: Creando reglas de asociación

Una vez aplicado el algoritmo y obtenidas las reglas, en las siguientes secciones se van a filtrar las reglas para obtener aquellas que sean más interesantes.

3.1. Reglas que más veces ocurren

En la Listing 13 se muestra el código usado para obtener las 10 reglas que más veces ocurren.

```
1 inspect(sort(rules, by='count')[1:10])
```

Listing 13: Reglas que más veces ocurren

Por otro lado, en la Figura 8 se muestran las 10 reglas que más veces ocurren.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.8716	0.9641593	0.9040	1.026138	4358
[2]	{Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.8646	0.9653863	0.8956	1.027444	4323
[3]	{Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.7984	0.9837358	0.8116	1.046973	3992
[4]	{CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.6936	0.9824363	0.7060	1.045590	3468
[5]	{Personal Loan=No acepto prestamo personal, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.6356	0.9953022	0.6386	1.059283	3178
[6]	{Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.6276	0.9958743	0.6302	1.059892	3138
[7]	{Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.6094	0.9673016	0.6300	1.029482	3047
[8]	{Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.6022	0.9716037	0.6198	1.034061	3011
[9]	{CCAvg=[0-2]}	=> {Personal Loan=No acepto prestamo personal}	0.5934	0.9699248	0.6118	1.072926	2967
[10]	{CCAvg=[0-2]}	=> {CD Account=No tiene cuenta CD}	0.5878	0.9607715	0.6118	1.022532	2939

Figura 8: Las 10 primeras reglas que más veces ocurren

La regla que más veces ocurre con una confianza del 96 % es: “Si el cliente no aceptó el préstamo personal el año pasado, entonces no tiene certificado de depósito”.

3.2. Reglas cuyo RHS = Securities Account

Uno de los productos que ofrece el banco son las cuentas de valores. En la Listing 15 se muestra el código que se ha usado para filtrar aquellas reglas cuyo RHS sea **Securities Account**. Además, se ha filtrado para que se muestren las 10 primeras reglas ordenadas por la confianza y por el número de instancias que cumplen la regla.

```
1 # Subset RHS = Securities Account
2 rules.securities_account <- subset(rules, subset = rhs %pin% "Securities Account=")
3 inspect(sort(rules.securities_account, by='confidence')[1:10])
4 inspect(sort(rules.securities_account, by='count')[1:10])
```

Listing 14: Reglas cuyo RHS es **Securities Account**

En la Figura 9 se muestran las reglas de asociación encontradas. Hay dos reglas que ocurren siempre:

1. Si el usuario no tiene cuenta de certificado de depósito, usa la banca online y usa tarjeta de crédito, entonces no tiene cuenta de valor.
2. Si el usuario no aceptó el préstamo personal el año pasado, no tiene cuenta de certificado de depósito, usa la banca online y usa tarjeta de crédito, entonces no tiene cuenta de valor.

Por otro lado, hay 2 reglas que no llegan al 100% de confianza pero que ocurren un número mayor de veces.

1. Si el usuario no tiene cuenta de certificado de depósito y usa tarjeta de crédito, entonces no tiene cuenta de valor.
2. Si el usuario no aceptó el préstamo personal el año pasado, no tiene cuenta de certificado de depósito y usa tarjeta de crédito, entonces no tiene cuenta de seguridad.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{CD Account=No tiene cuenta CD, Online=Usa online, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.1316	1.0000000	0.1316	1.116570	658
[2]	{Personal Loan=No acepto prestamo personal, CD Account=No tiene cuenta CD, Online=Usa online, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.1316	1.0000000	0.1316	1.116570	658
[3]	{Education=Education level, CD Account=No tiene cuenta CD, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.1040	0.9701493	0.1072	1.083239	520
[4]	{Education=Education level, Personal Loan=No acepto prestamo personal, CD Account=No tiene cuenta CD, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.1018	0.9695238	0.1050	1.082541	509
[5]	{CD Account=No tiene cuenta CD, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.2370	0.9634146	0.2460	1.075720	1185
[6]	{Personal Loan=No acepto prestamo personal, CD Account=No tiene cuenta CD, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.2270	0.9618644	0.2360	1.073989	1135
[7]	{CCAvg=[0-2], CD Account=No tiene cuenta CD, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.1558	0.9593596	0.1624	1.071192	779
[8]	{CCAvg=[0-2], Personal Loan=No acepto prestamo personal, CD Account=No tiene cuenta CD, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.1542	0.9589552	0.1608	1.070741	771
[9]	{Mortgage=No tiene hipoteca, CD Account=No tiene cuenta CD, CreditCard=Usa tarjeta}	=> {Securities Account=No tiene cuentas de seguridad}	0.1670	0.9531963	0.1752	1.064310	835
[10]	{Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, CD Account=No tiene cuenta CD,						

Figura 9: Las 10 primeras reglas con más confianza con **Securities Account**

3.3. Reglas cuyo RHS = CD Account

Otro producto del banco son las cuentas de certificado de depósito. En la Listing 13 se muestra el código que se ha usado para filtrar estas reglas. Además, se ha filtrado para que se muestren las 10 primeras reglas ordenadas por la confianza y el número de instancias que cumplen la regla.

```

1  # Subset RHS = CD Account
2  rules.cd_account <- subset(rules, subset = rhs %pin% "CD Account=")
3  inspect(sort(rules.cd_account, by='confidence')[1:10])
4  inspect(sort(rules.cd_account, by='count')[1:10])

```

Listing 15: Reglas cuyo RHS es **CD Account**

En la Figura 10 se muestran las 10 reglas de asociación que tienen mayor confianza. Se puede ver que todas tienen una confianza del 100%, con lo cual se cumplen siempre.

Por otro lado, en la Figura 11 se muestran las 10 reglas que más veces ocurren.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Income=[33-52], Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1224	1	0.1224	1.064283	612
[2]	{Income=[8-33], Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1226	1	0.1226	1.064283	613
[3]	{Income=[52-78], Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1290	1	0.1290	1.064283	645
[4]	{CCAvg=[0-2], Mortgage=Hipoteca normal, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1208	1	0.1208	1.064283	604
[5]	{Family=2, Securities Account=No tiene cuentas de seguridad, Online=No usa online}	=> {CD Account=No tiene cuenta CD}	0.1000	1	0.1000	1.064283	500
[6]	{Education=Education level, Securities Account=No tiene cuentas de seguridad, Online=No usa online}	=> {CD Account=No tiene cuenta CD}	0.1516	1	0.1516	1.064283	758
[7]	{CCAvg=[0-2], Online=No usa online, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1732	1	0.1732	1.064283	866
[8]	{CCAvg=[0-2], Securities Account=No tiene cuentas de seguridad, Online=No usa online}	=> {CD Account=No tiene cuenta CD}	0.2222	1	0.2222	1.064283	1111
[9]	{Securities Account=No tiene cuentas de seguridad, Online=No usa online, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.2568	1	0.2568	1.064283	1284
[10]	{Personal Loan=No acepto prestamo personal, Online=No usa online, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.2600	1	0.2600	1.064283	1300

Figura 10: Las 10 primeras reglas con más confianza con **CD Account**

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.8716	0.9641593	0.9040	1.026138	4358
[2]	{Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.8646	0.9653863	0.8956	1.027444	4323
[3]	{Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.7984	0.9837358	0.8116	1.046973	3992
[4]	{CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.6936	0.9824363	0.7060	1.045590	3468
[5]	{Personal Loan=No acepto prestamo personal, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.6356	0.9953022	0.6386	1.059283	3178
[6]	{Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.6276	0.9958743	0.6302	1.059892	3138
[7]	{Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.6094	0.9673016	0.6300	1.029482	3047
[8]	{Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.6022	0.9716037	0.6198	1.034061	3011
[9]	{CCAvg=[0-2]}	=> {CD Account=No tiene cuenta CD}	0.5878	0.9607715	0.6118	1.022532	2939
[10]	{CCAvg=[0-2], Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.5744	0.9679811	0.5934	1.030206	2872

Figura 11: Las 10 primeras reglas que más veces ocurren con **CD Account**

Destaco una regla que ocurre muchas veces que es: “Si la media de gastar dinero por mes está entre 0-2, entonces el cliente no tiene certificados de depósito”. Los certificados de depósito los suelen tener clientes que gastan más dinero a lo largo del mes.

3.4. Reglas cuyo LHS = Income y Mortgage

Sería interesante saber qué reglas son las que están relacionadas con el salario anual e hipoteca de los clientes. Para ello, se ha usado el código de la Listing 16.

```

1 # Subset LHS = Income y Mortgage
2 rules.income <- subset(rules, subset = lhs %pin% "Income=" & lhs %pin% "Mortgage=")
3 inspect(sort(rules.income, by='confidence')[1:10])
4 inspect(sort(rules.income, by='count')[1:10])

```

Listing 16: Reglas cuyo LHS es **Income** y **Mortgage**

En la Figura 12 se muestran las 10 reglas de asociación que tienen mayor confianza.

Por otro lado, en la Figura 13 se muestran las 10 reglas que más veces ocurren.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Income=[33-52), Mortgage=No tiene hipoteca}	=> {Personal Loan=No acepto prestamo personal}	0.1360	1	0.1360	1.106195	680
[2]	{Income=[8-33), Mortgage=No tiene hipoteca}	=> {Personal Loan=No acepto prestamo personal}	0.1320	1	0.1320	1.106195	660
[3]	{Income=[33-52), CCAvg=[0-2), Mortgage=No tiene hipoteca}	=> {Personal Loan=No acepto prestamo personal}	0.1032	1	0.1032	1.106195	516
[4]	{Income=[33-52), Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {Personal Loan=No acepto prestamo personal}	0.1204	1	0.1204	1.106195	602
[5]	{Income=[33-52), Mortgage=No tiene hipoteca, CD Account=No tiene cuenta CD}	=> {Personal Loan=No acepto prestamo personal}	0.1312	1	0.1312	1.106195	656
[6]	{Income=[8-33), CCAvg=[0-2), Mortgage=No tiene hipoteca}	=> {Personal Loan=No acepto prestamo personal}	0.1300	1	0.1300	1.106195	650
[7]	{Income=[8-33), Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {Personal Loan=No acepto prestamo personal}	0.1194	1	0.1194	1.106195	597
[8]	{Income=[8-33), Mortgage=No tiene hipoteca, CD Account=No tiene cuenta CD}	=> {Personal Loan=No acepto prestamo personal}	0.1272	1	0.1272	1.106195	636
[9]	{Income=[33-52), Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad, CD Account=No tiene cuenta CD}	=> {Personal Loan=No acepto prestamo personal}	0.1188	1	0.1188	1.106195	594
[10]	{Income=[8-33), CCAvg=[0-2), Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {Personal Loan=No acepto prestamo personal}	0.1174	1	0.1174	1.106195	587

Figura 12: Las 10 primeras reglas con más confianza con **Income** y **Mortgage**

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Income=[78-113], Mortgage=No tiene hipoteca}	=> {CD Account=No tiene cuenta CD}	0.1392	0.9586777	0.1452	1.020304	696
[2]	{Income=[52-78], Mortgage=No tiene hipoteca}	=> {Personal Loan=No acepto prestamo personal}	0.1376	0.9956585	0.1382	1.101392	688
[3]	{Income=[33-52], Mortgage=No tiene hipoteca}	=> {Personal Loan=No acepto prestamo personal}	0.1360	1.0000000	0.1360	1.106195	680
[4]	{Income=[52-78], Mortgage=No tiene hipoteca}	=> {CD Account=No tiene cuenta CD}	0.1346	0.9739508	0.1382	1.036559	673
[5]	{Income=[52-78], Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.1342	0.9752907	0.1376	1.037985	671
[6]	{Income=[52-78], Mortgage=No tiene hipoteca, CD Account=No tiene cuenta CD}	=> {Personal Loan=No acepto prestamo personal}	0.1342	0.9970282	0.1346	1.102907	671
[7]	{Income=[8-33], Mortgage=No tiene hipoteca}	=> {Personal Loan=No acepto prestamo personal}	0.1320	1.0000000	0.1320	1.106195	660
[8]	{Income=[78-113], Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.1314	0.9690265	0.1356	1.031318	657
[9]	{Income=[33-52], Mortgage=No tiene hipoteca}	=> {CD Account=No tiene cuenta CD}	0.1312	0.9647059	0.1360	1.026720	656
[10]	{Income=[33-52], Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.1312	0.9647059	0.1360	1.026720	656

Figura 13: Las 10 primeras reglas que más veces ocurren con **Income** y **Mortgage**

3.5. Reglas cuyo LHS = Age y Mortgage

También resulta interesante ver la relación que hay entre la edad de los clientes y la hipoteca. Estas reglas se pueden ver en el código de la Listing 17.

```

1 # Subset LHS = Age y Mortgage
2 rules.education <- subset(rules, subset = lhs %pin% "Age=" & lhs %pin% "Mortgage")
3 inspect(sort(rules.education, by='confidence')[1:10])
4 inspect(sort(rules.education, by='count')[1:10])

```

Listing 17: Reglas cuyo LHS es **Age** y **Mortgage**

En la Figura 14 se muestran las 7 reglas de asociación que tienen mayor confianza. En realidad son 10 las que se muestran, pero no caben en la captura de pantalla. Las 10 reglas en total pueden verse en el código que se ha adjuntado en esta práctica.

Por otro lado, en la Figura 15 se muestran las 10 reglas que más veces ocurren.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Age=45-54, Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1060	1.0000000	0.1060	1.064283	530
[2]	{Age=35-44, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1040	1.0000000	0.1040	1.064283	520
[3]	{Age=55-67, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1102	1.0000000	0.1102	1.064283	551
[4]	{Age=35-44, Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1124	0.9982238	0.1126	1.062392	562
[5]	{Age=55-67, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1542	0.9961240	0.1548	1.060158	771
[6]	{Age=55-67, CCAvg=[0-2), Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1014	0.9960707	0.1018	1.060101	507
[7]	{Age=55-67, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, CreditCard=No usa tarjeta}	=> {CD Account=No tiene cuenta CD}	0.1210	0.9950650	0.1216	1.059031	605

Figura 14: Las 10 primeras reglas con más confianza con **Age** y **Mortgage**

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Age=55-67, Mortgage=No tiene hipoteca}	=> {CD Account=No tiene cuenta CD}	0.1790	0.9521277	0.1880	1.013333	895
[2]	{Age=55-67, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.1678	0.9721900	0.1726	1.034685	839
[3]	{Age=55-67, Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1652	0.9833333	0.1680	1.046545	826
[4]	{Age=35-44, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.1574	0.9728059	0.1618	1.035340	787
[5]	{Age=55-67, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1542	0.9961240	0.1548	1.060158	771
[6]	{Age=45-54, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal}	=> {CD Account=No tiene cuenta CD}	0.1526	0.9609572	0.1588	1.022730	763
[7]	{Age=35-44, Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1524	0.9657795	0.1578	1.027862	762
[8]	{Age=45-54, Mortgage=No tiene hipoteca, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1508	0.9654289	0.1562	1.027489	754
[9]	{Age=35-44, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1430	0.9862069	0.1450	1.049603	715
[10]	{Age=45-54, Mortgage=No tiene hipoteca, Personal Loan=No acepto prestamo personal, Securities Account=No tiene cuentas de seguridad}	=> {CD Account=No tiene cuenta CD}	0.1396	0.9803371	0.1424	1.043356	698

Figura 15: Las 10 primeras reglas que más veces ocurren con **Age** y **Mortgage**

En estas reglas de asociación se puede ver que el rango de edad [23-35) años no aparece en ninguna regla, y que el rango de edad mayoritario es el de [55-67] años. Además, la consecuencia de todas estas reglas es que no tienen certificados de depósito.